

Knowledge encapsulation : studies on the development of medical expertise

Citation for published version (APA):

van de Wiel, M. W. J. (1997). *Knowledge encapsulation : studies on the development of medical expertise*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht.
<https://doi.org/10.26481/dis.19970912mw>

Document status and date:

Published: 01/01/1997

DOI:

[10.26481/dis.19970912mw](https://doi.org/10.26481/dis.19970912mw)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Knowledge encapsulation

Studies on the development of medical expertise

Knowledge encapsulation

Studies on the development of medical expertise

Proefschrift

Ter verkrijging van de graad van doctor aan
de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. mr. M. J. Cohen
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 12 september 1997 om 16.00 uur

door

Margaretha Wilhelmina Johanna van de Wiel

Promotor:

Prof. dr. H. G. Schmidt

Co-promotor:

Dr. H. P. A. Boshuizen

Beoordelingscommissie:

Prof. dr. C. P. M. van der Vleuten (voorzitter)

Prof. dr. J. J. Elshout (Universiteit van Amsterdam)

Dr. A. Landeweerd

Prof. dr. H. L. G. J. Merckelbach

Prof. dr. A. J. W. Thomassen (Katholieke Universiteit Nijmegen)

Omslag: M. G. Hoevenaars, Alatis Patent, Eindhoven

Druk: Ponsen & Looijen, Wageningen

ISBN 90-9010793-2

Copyright © M. W. J. van de Wiel, 1997

All rights reserved. No part of this book may be reproduced in any form or by any means, electronical or mechanical, including photocopying, recording or by any information storage and retrieval system without permission of the author.

Voor mijn oma,
vader en moeder

Contents

Preface	xi
1 The development of expert medical knowledge: Three perspectives	1
A historical view on the different aspects of medical knowledge	1
The role of biomedical knowledge in clinical reasoning	3
A central role of biomedical knowledge	4
Biomedical and clinical knowledge as two worlds apart	5
Encapsulation of biomedical knowledge into clinical knowledge	7
Conclusion	11
Mechanisms of learning	12
Overview of the present studies	14
References	16
2 Robustness of the intermediate effect in clinical case recall	19
Method	29
Results	31
Incidental versus intentional recall condition	31
Numerical versus interpreted lab data condition	33
Discussion	35
References	38
Appendix	40
3 Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians	41
Method	51
Results	56

Diagnostic accuracy	56
Pathophysiological explanations	57
Discussion	63
References	69
Appendices	70
4 The explanation of clinical concepts by expert physicians, clerks and advanced students	75
Method	79
Results and discussion	82
Elaborateness of explanations	82
Quality of explanations	83
Fluency of explanations	84
Qualitative analysis of the results	87
Validity of the quality measurement	89
General discussion	90
References	92
Appendix	93
5 A failure to reproduce the intermediate effect in clinical case recall	95
Method	99
Results	100
Discussion	102
References	105
Appendix	107
6 Knowledge encapsulation revisited: Summary and general discussion	109
Knowledge encapsulation	110
Summary of the present studies	113
The influence of recall instructions and lab data format on clinical case recall	113
Pathophysiological representations of clinical cases	114
Explanation of encapsulating concepts	116
Clinical case recall revisited	117
Reformulation of Schmidt and Boshuizen's theory on expertise development in medicine	120
The development of the knowledge network from student to expert	122
Diagnosis of routine clinical cases by students and experts	124
Non-routine diagnosis and explanation by experts	126

Future research	130
Implications for medical education	131
References	134
Appendices	137
Samenvatting	139
Curriculum Vitae	143

Preface

What happens in that black box of the brain while thinking? What is knowledge? And how does expertise develop? These questions inspired me to apply for a position as Ph.D. researcher at the Department of Educational Development and Research in Maastricht in the domain of medical expertise. After writing this thesis I am still happy I did.

When I started my research in 1991, Henk Schmidt and Els Boshuizen, my promotor and co-promotor, had already done substantial research in the domain, and had formulated some testable notions on how medical expertise develops. It became my job to test their notion of knowledge encapsulation. The enthusiasm and curiosity of all three of us has resulted in a series of studies. Four of these studies are presented in this thesis.

Writing a thesis is teamwork. I would like to thank my team members, Henk and Els, for their good ideas, their critical comments and the lively discussions we had.

There was a larger group of researchers on expertise in Maastricht that I would like to thank. With Rina Vaatstra, Remy Rikers, Eugene Custers and Herco Fonteijn I enjoyed exchanging thoughts that were in any way related to our research.

Furthermore, I am grateful I had the possibility to learn from colleagues abroad. First of all, I would like to thank the group at the Centre for Medical Education in Montreal. In particular, Vimla Patel, Dave Kaufman and José Arocha sharpened my mind by challenging the Maastricht idea of knowledge encapsulation. Asun Lopez-Manjon became a dear friend. I thank Arthur Elstein for discussing my work in Chicago, which was a real pleasure. I also want to mention Paul Feltovich and Geoff Norman.

A group of people I am greatly indebted to are the expert physicians who helped me to construct the case materials and model explanations used in the studies. These are Stefan Bakker, Robert-Jan Brummer, Gerard Essed, Hans Fiolet, Pie Hobus, Harry Hulsmans, Peter de Leeuw, Wubbo Mulder, and Nicolaas Schaper. I owe special thanks to Nicolaas, who has been my

personal tutor in internal medicine for several years now. His creative ideas and efficiency meant that his contribution went far beyond his own domain.

I appreciate that so many clinicians and students were willing to participate in the studies. Without them this thesis would have been impossible. In particular, I am indebted to the internists at the university hospital in Maastricht (AZM) to whom I appealed more than once.

I thank Maaïke de Vries for being my extra hand in collecting, transcribing and analyzing data for many years. Other research assistants I am grateful to are Ingrid Hertogs, Ceciel van Hemel, Eline Tiems and Karen de Heus.

I thank my colleagues at the Department of Educational Development and Research for their interest and advice. My special thanks go to my roommates and Cita van Til, who shared the daily stress. I thank Albert Scherpbier and the Skillslab for providing a stimulating work environment to complete the job. I thank Bob Wilkinson for carefully reading and correcting the text.

Finally, I would like to thank my family and friends, who contributed each in their own way to the realization of this thesis. I mention some. Mark Hoevenaars designed the professional cover. Pon Pennings listened to me so often. My grandma reminded me to take a break. My parents have always given me their full support. Chris Ingelse took care of me during the last few years when our patience has been tried.

Margje van de Wiel
Maastricht, July 1997

The development of expert medical knowledge:

Three perspectives

A historical view on the different aspects of medical knowledge

Disease and death are so inherent to life that human beings in all times have been puzzled by the origins and treatment of disease. The medicinal value of some herbs was discovered long ago by trial and error in early folk medicine. These were used to remedy common maladies, such as cold and constipation. More serious illnesses, however, were thought to be of supernatural origin, and magic and religious rites were used to overcome or expel the evil. It was in the 5th century BC in Greece that disease came to be regarded as a natural rather than a supernatural phenomenon. According to the doctrine of the four bodily humors, health was maintained by the harmony of blood, phlegm, yellow bile and black bile, whereas disease was a consequence of a bad mix of these four humors. The works of Hippocrates represented these ideas. These works, moreover, emphasized the importance of careful observation of patients to predict the course of disease, as well as the importance of the study of patients' environments to find clues for external causes of disease. The Hippocratic method and theory was followed and extended by Galen (130-200 AD), a physician in Rome. He studied the anatomy and physiology of animals by means of dissection and experimentation. Galen developed a theory on the motion of the blood in which the doctrine of the humors was integrated. This theory remained for centuries the medical authority in Europe. It was not until Vesalius had published his work on the anatomy of the human body (1543), and Harvey had published his work on the physiology of the circulation of the blood (1628), that Galen's theory was seriously disputed. In the meantime, the Christian Church provided medical care and spiritual assistance, whereas surgeons dealt with the practical treatment of disease, such as repairing fractures and dislocations, performing amputations, and treating wounds, and midwives took care of childbirth. Thus, medical practice relied heavily on experience

with diseases and their treatment on the one hand, and on the prevailing theoretical explanations on the other.

Since the foundation of the first medical schools in the middle ages, theory was taught at universities. Students spent most of their time following lectures and reading medical texts. Sometimes they could also attend demonstrations of the dissection of corpses. However, practical experience with patients was usually only acquired after students had completed the theoretical part, and worked as an apprentice or a physician. From the 16th century onwards, a few schools provided students, in addition to theoretical lessons, with clinical lessons at the bedside allowing them to observe the manifestations of disease described in the literature. Exactly the opposite happened in the training of surgeons and midwives, who started as apprentices and learned their profession first by seeing and doing. Only in a later phase of training did they follow some courses organized by their professional guilds.

As a result of scientific inquiries in the 17th and 18th century, the theories and subjects taught at the universities gradually changed. Sciences such as physics, chemistry and botany became part of the medical curriculum, as well as anatomy and physiology. These sciences were considered to be the theoretical basis of medicine and, hence, called the basic sciences. In the 19th century, the body of biomedical knowledge, i.e. the knowledge contributing to the understanding of the functioning and dysfunctioning of the human body, grew exponentially, and the methods available for examining and treating patients steadily increased and improved. The development of sciences such as cellular pathology, histology, virology and bacteriology had a great part in providing explanations for disease that were not merely theoretical, but scientifically proven. In this way, a sound causal relationship was established between the biomedical sciences and the clinical sciences concerned with the manifestations, classification and treatment of disease. This development had important consequences for both medical practice and education. In the mid-19th century, the clinical practices of surgeons and midwives came to an end, when in many European countries and the United States, all medical practitioners needed to have a registration as qualified physician by the national medical board. Surgery and obstetrics became incorporated in the medical curriculum, and national examinations were introduced. By 1900 it was widely acknowledged that the basic sciences constituted the foundation on which clinical knowledge should be built. A highly influential representative of this vision was Abraham Flexner (Bordley & Harvey, 1976; Duffy, 1993). At the instigation of the Council of Medical Education of the American Medical Association he published a report (Flexner, 1910) in which he analyzed the different curricula of North American medical schools and formulated recommendations for reforms. He advocated a four-year medical curriculum consisting of two years of basic

science teaching¹ followed by two years of clinical practice in the hospital.² The scientific nature of the training would be guaranteed by giving medical researchers the responsibility for teaching medical students in the preclinical years.

For most of this century the traditional curriculum model with a pre-clinical and clinical phase has dominated medical education. However, since the seventies the adequacy of this model has become subject to considerable debate (e.g., Balla, Biggs, Gibson & Chang, 1990; Barrows & Tamblyn, 1980; Bouman & Snellen-Balendong, 1996; Cavazos, 1984; Maddison, 1978; Neame, 1984; Patel, Evans & Groen, 1989b; Patel & Kaufman, 1995; Schmidt & De Volder, 1984; Simpson, 1972). The traditional curriculum was criticized in that the basic science programs were too scientifically oriented and had lost sight of their primary objective of equipping students with the knowledge necessary to underpin the development of clinical understanding and competence. It was not doubted that biomedical knowledge played a crucial role in medical education, but rather in what way and to what extent it should be provided (e.g., Barrows & Tamblyn, 1980; Dawson-Saunders, Feltovich, Coulson & Steward, 1990; Neame, 1984; Schmidt & De Volder, 1984). The progressive growth of the body of biomedical knowledge, the development of additional clinical specializations, and the need for training of clinical skills and attitudes, furthermore, contributed to the practical problem of how to schedule all disciplines in the available time. This gave rise to the more fundamental question of what role each of the disciplines plays in the development of good medical practitioners.

The present thesis studies the role of biomedical knowledge in understanding and diagnosing clinical problems from a cognitive psychological point of view. Three distinct views on this topic will be discussed in this chapter. Before doing so, the rationale and method of cognitive psychological experimentation in research on medical expertise will be briefly explained.

The role of biomedical knowledge in clinical reasoning

Two decades ago, cognitive psychologists started to investigate the nature of medical expertise. Initially, research was focused on the problem solving processes of medical students and expert physicians in clinical diagnosis (Elstein, Shulman & Sprafka, 1978; Neufeld, Norman, Feightner & Barrows,

¹ The term "basic sciences" is, in fact, somewhat confusing, because the sciences considered to be the basis of medicine changed over time and among medical schools. The basic science program advocated by Flexner (1910) consisted of biomedical sciences, such as anatomy, physiology, pharmacology and pathology. Education in the sciences of chemistry, physics, and biology was set as requirements for admission to medical school.

² This curriculum model is known as the conventional or traditional medical curriculum.

1981). These early studies showed that both students and experts approached the diagnostic task by first generating diagnostic hypotheses and then seeking additional information to confirm or refute the hypotheses. Experts, however, generated much more accurate diagnostic hypotheses than students, often already at an early stage in the patient encounter. Medical expertise thus appeared to be primarily based on an adequate body of domain-specific knowledge. Therefore, research focused on the nature and organization of expert medical knowledge, and the changes that occur in the knowledge base during the development from novice to expert. The basic paradigm applied in this research is to compare the knowledge subjects of different levels of expertise use in diagnosing and representing clinical cases. A clinical case usually consists of a description of a patient's complaints and symptoms, the results of physical examination, the relevant laboratory data and the results of additional examinations. In think-aloud studies of diagnostic reasoning subjects are asked to verbalize whatever comes up in their minds in proceeding through a case. This method, thus, provides an on-line protocol of subjects' thinking (Elstein et al., 1978; Ericsson & Simon, 1984). In clinical case representation studies subjects are asked to study a case in order to provide a diagnosis, to recall whatever they remember from the case, and/or to provide a pathophysiological explanation for the signs and symptoms in the case. Clinical case recall is assumed to reflect the content and structure of the representation formed in case processing (Muzzin et al., 1982; Norman, Jacoby, Feightner & Campbell, 1979; Patel & Groen, 1986), and pathophysiological case explanations are assumed to reflect the causal knowledge underlying case representations (Feltovich & Barrows, 1984; Patel & Groen, 1986).

The results of these types of studies, however, led to three different views on the role that biomedical knowledge plays in expert clinical reasoning, and, hence, in developing expert medical knowledge.

A central role of biomedical knowledge

In the early 1980s various researchers assumed that biomedical knowledge fulfills a central and integrating function in medical diagnosis (Feltovich & Barrows, 1984; Feltovich, Johnson, Moller & Swanson, 1984; Kuipers & Kassirer, 1984; Lesgold et al., 1988; Lesgold, 1984). These researchers argued that the key to correct diagnostic thinking is the deep understanding of the normal anatomy and physiology of the human body. Medical students and doctors were assumed to interpret deviations from the normal structure and functioning in a patient by reconstructing the cause and course of disease in that patient. Elaborate knowledge of the pathophysiology of diseases and of contextual factors influencing the process of disease in particular patients would be needed to make a clinical diagnosis. According to these authors, the integral role of biomedical knowledge is to constrain and guide the clinical reasoning process. In their view, a lack in deep conceptual un-

derstanding of the biomedical knowledge underlying a disease would result in diagnostic errors.

This conceptualization of the role of biomedical knowledge in clinical reasoning was empirically supported by several investigations. Lesgold and colleagues (1984; 1988) performed a set of experiments on the diagnosis of x-ray pictures in radiology. The three groups of subjects participating in these studies were first- and second-year residents in radiology, third- and fourth-year residents in radiology, and expert radiologists. Subjects were asked to think aloud while examining x-ray films in order to formulate a diagnosis. It appeared that the experts and residents differed in the precision with which they analyzed the films. The experts described the abnormalities in the films in anatomically more specific terms. They also had a greater ability to recognize (overlapping) anatomical structures and to interpret film features in the light of a particular patient and as the result of the imaging technique. For instance, experts concluded that the heart size was within normal limits taking into account the patient's curvature in the spine, or the patient's poor inspiration. The think-aloud protocols of experts, furthermore, were more extensive than those of residents and showed more connections between the reported findings. The latter finding was interpreted as a higher level of inferential thinking in experts. Lesgold and colleagues concluded that expertise in radiology is determined by the capability to envision a patient's internal anatomy in examining x-ray pictures, to which end detailed knowledge of normal and pathological anatomy is required.

In an experiment by Feltovich et al. (1984), subjects with three levels of experience in pediatric cardiology were asked to diagnose four paper cases of congenital heart disease while thinking aloud. Each case was selected to assess a different aspect of disease knowledge. The results showed that more experienced subjects considered more adequate sets of relevant hypotheses which were evaluated on the basis of their underlying pathophysiological knowledge of congenital heart disease. In more recent work, Feltovich and colleagues (Feltovich, Spiro & Coulson, 1989; Feltovich, Coulson, Spiro & Dawson-Saunders, 1992; Feltovich, Spiro & Coulson, 1993) focused on the nature of misconceptions prevalent in the understanding of complex biomedical systems. They argued that the failure of deep conceptual understanding in biomedicine contributes to inadequacies in the application of knowledge in clinical tasks. It is the opinion of these researchers that establishing detailed understanding of the normal and abnormal functioning of the human body will help students to develop knowledge that can be flexibly applied in and transferred among clinical situations.

Biomedical and clinical knowledge as two worlds apart

The perspective that reasoning with biomedical knowledge would be necessary to generate an accurate clinical diagnosis was challenged by Patel, Groen

and colleagues (Patel, Evans & Groen, 1989a; Patel et al., 1989b; Patel & Kaufman, 1995), who suggested that the biomedical sciences and the clinical sciences constitute two different worlds, with their own modes of reasoning. These researchers rejected the assumption that medical practitioners develop one coherent medical knowledge base, and hypothesized that knowledge of basic sciences and clinical medicine would rather be organized in two separate knowledge bases with distinctly different knowledge structures. In their view, clinical knowledge would be based on a complex taxonomy relating symptoms to disease categories. Biomedical knowledge, on the other hand, would be based on general principles defining chains of causal mechanisms. They argued that clinical knowledge would be used in solving medical problems, whereas biomedical knowledge would be used to provide coherence in the explanation and communication of biomedical and clinical phenomena (Patel et al., 1989b; Patel & Kaufman, 1995).

The experimental data which have led Patel et al. (1989a; 1989b; Patel & Kaufman, 1995) to the two-world hypothesis can be summarized as "the lack of basic science concepts in the reasoning about medical problems". This finding was first observed by Patel and Groen in their 1986 study on the reasoning strategies of experts in medical problem solving. Cardiologists were asked to read a description of a case of bacterial endocarditis, to recall the information in the case, to describe the underlying pathophysiology of the case, and to formulate a diagnosis. The diagnostic explanations showed that experts predominantly reasoned in a bottom-up way from clinical data to diagnostic hypotheses and sometimes in a top-down way from diagnostic hypotheses to clinical data. Basic science knowledge was not used in these patterns of reasoning which were both based on clinical associations.

In a subsequent study, Patel, Groen and Arocha (1990) investigated the diagnostic reasoning of physicians being either practitioners or researchers. Subjects were requested to read, recall, explain and diagnose two cases that included symptoms indicative for another disease than the main diagnosis. The diagnostic explanations showed that the practitioners largely relied on direct associations between clinical data and diagnostic hypotheses. These direct associations were used to describe the patient's condition, as well as to rule out alternative diagnoses. Pathophysiological knowledge was only used to account for the signs and symptoms that were not consistent with the hypothesis being considered. The medical researchers, on the other hand, largely explained the cases in basic pathophysiological terms. They rather represented the biomedical functioning of a patient, and did not even attempt to use their biomedical knowledge to develop a clinical description of the patient. Patel et al. (1989a; 1989b) interpreted these observations as support for the notion that basic science does not contribute directly to reasoning in clinical problem solving.

Furthermore, two experiments were performed to elucidate the role of basic science knowledge in clinical reasoning by medical students (Patel,

Evans & Kaufman, 1990; Patel, Groen & Scott, 1988). In addition to a clinical case of acute bacterial endocarditis, subjects were presented with three basic science texts relevant to this case. In the study by Patel, Groen and Scott, the basic science texts were presented before the clinical case, whereas in the study by Patel, Evans and Kaufman, the basic science texts were presented after the clinical case was diagnosed. In both studies subjects were first-, second- and fourth-year medical students. In the first study (Patel et al., 1988) subjects read and recalled basic science texts on the physiology of fever, hemodynamics and microcirculation at a first session, and read, recalled, diagnosed and explained the case of acute bacterial endocarditis at a second session. The results showed that all three groups of subjects recalled more information from the basic science texts than they applied in their explanations. Only a small proportion of the explanation protocols was based on the presented basic science information. According to the authors, this might indicate that the students had difficulty in relating the scientific information to the clinical problem. The knowledge actually used in explaining the clinical case differed for the three subject groups. The first-year students mainly used common-sense experiential knowledge. The second-year students made extensive use of biomedical knowledge other than that in the basic science texts, but the inferences they generated were frequently inconsistent and incorrect. The fourth-year students, who already had clinical experience as clerks, more often correctly combined clinical and biomedical concepts to construct causal explanations for the findings.

In the second study (Patel, Evans & Kaufman, 1990) subjects read, summarized, explained and diagnosed the case of bacterial endocarditis. Then they read the three basic science texts and were asked to explain the clinical problem again. The results showed that only the fourth-year students who had already developed a strong diagnostic hypothesis could use the information in the basic science texts to improve their diagnoses by explaining the relations between the clinical findings. Therefore, the authors concluded that the effectiveness of the application of biomedical knowledge for diagnostic reasoning depends on the availability of disease classification knowledge. The results of both of these experiments were interpreted by Patel et al. (1988; 1989a; Patel, Evans & Kaufman, 1990) as support for their hypothesis that diagnostic problem solving is initially a matter of direct clinical association, and that biomedical knowledge only comes into play to explain the relations between findings when clinical knowledge falls short.

Encapsulation of biomedical knowledge into clinical knowledge

The theoretical notion of knowledge encapsulation introduced by Schmidt and Boshuizen (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992, 1993a) takes a position in the debate between the two extremes sketched so far. This position does not assume that biomedical knowledge fulfills a central function in the diagnostic reasoning process itself, neither does it as-

sume that biomedical knowledge only has a function in the explanation of clinical phenomena, but it rather assumes that biomedical knowledge has an important function in the development of easily applicable clinical knowledge. Schmidt and Boshuizen (1992) defined knowledge encapsulation as the subsumption or packaging of "lower level" detailed concepts and their interrelations in an associative net, under a smaller number of "higher level" concepts with the same explanatory power. They assumed that through repeated application of knowledge in medical training and practice, elaborate networks of causal biomedical knowledge become encapsulated into more comprehensive clinical concepts or diagnostic labels. Therefore, complex pathophysiological processes relating causes and consequences of disease could be captured in a few encapsulating concepts. In this way, causal biomedical knowledge would become incorporated into clinical knowledge that was assumed to be represented in list-like structures called illness scripts. Through practical experience in the clinic these illness scripts would become more and more tuned towards the cases encountered. In diagnosing routine cases, therefore, experts would use highly tuned clinical knowledge. But, experts would also be able to unfold encapsulated causal biomedical knowledge, whenever required for task performance.

Boshuizen and Schmidt (1992) argued that Patel et al.'s finding (Patel, Evans & Kaufman, 1990; Patel & Groen, 1986; Patel, Groen & Arocha, 1990; Patel et al., 1988) that clinical rather than biomedical concepts are applied by expert physicians in medical problem solving, could also be explained as a result of knowledge encapsulation. They performed two studies in which they were able to reproduce this finding of Patel and colleagues, as well as to provide additional support for the notion that biomedical knowledge would become encapsulated and integrated into clinical knowledge. In the first study, a second-, fourth-, and fifth-year medical student, and a family physician were presented a case of pancreatitis. Subjects were instructed to think-aloud in proceeding through the case and to provide a differential diagnosis at the end. Analysis of the think-aloud protocols showed that the more experienced subjects used proportionally less biomedical concepts in diagnosing the case. The expert and fifth-year student almost entirely relied on clinical knowledge of disease, hardly using causal biomedical reasoning. The fourth-year student clearly used his biomedical knowledge to relate symptoms and to suggest diagnostic hypotheses. The second-year student elaborated extensively on his biomedical knowledge, but also made quite a few obvious errors.

In the second study, on-line clinical reasoning was combined with post-hoc explanations. Twenty subjects at the same four levels of expertise as in the first experiment participated. Subjects were asked to think-aloud while diagnosing the pancreatitis case, and then to describe the pathophysiological processes underlying the case. The think-aloud protocols showed that fourth-year students applied the highest proportion of biomedical knowl-

edge, whereas the family physicians applied the lowest proportion. The post-hoc pathophysiological explanation protocols, on the other hand, showed that the family physicians provided the most extensive explanations. This was interpreted as an indication that the lack of biomedical concepts in experts' protocols was not a result of decay of biomedical knowledge. In order to decide on the relation between the two data bases, Boshuizen and Schmidt analyzed the correspondences between the think-aloud protocols and the explanation protocols. This analysis revealed that the overlap between the two protocols increased with level of expertise: The clinical knowledge reasoned with in the think-aloud protocols appeared to be abbreviations of more detailed biomedical lines of reasoning in the post-hoc explanations. The authors concluded that biomedical knowledge is still available and accessible for experts, but plays a tacit role in diagnostic reasoning as it is encapsulated in clinical knowledge.

Furthermore, Schmidt and Boshuizen (1992; 1993a) claimed that their theory of knowledge encapsulation could explain two other important findings in medical expertise research: The general finding that experts produce accurate diagnoses in a much shorter time than students (e.g., Elstein et al., 1978; Neufeld et al., 1981; Norman, Rosenthal, Brooks, Allen & Muzzin, 1989), and the frequently observed finding that expert physicians represent clinical cases less elaborately than students of an intermediate level of expertise (Boshuizen, 1989; Patel & Medley-Mark, 1986; Schmidt, Boshuizen & Hobus, 1988). The latter finding is known as the *intermediate effect* in clinical case representation studies (Patel & Groen, 1991a; Patel & Groen, 1991b; Schmidt et al., 1988). Schmidt and Boshuizen argued that immediate activation of a relatively small number of comprehensive, encapsulating concepts in case processing allows experts to form quickly an adequate and condensed representation of the patient's problem. Because students have not yet developed knowledge in encapsulated form, and, hence, rely on earlier acquired biomedical knowledge, they would need more time than experts to construct a more or less coherent case representation. To test the hypothesis of knowledge encapsulation, Schmidt and Boshuizen (1993b) performed two clinical case representation studies in which they manipulated the time available for case processing. They expected that experts would process the case information in encapsulated form irrespective of the available processing time, whereas students would be impeded in elaborating on their knowledge if processing time was constrained.

In both studies, subjects at five different levels of expertise participated. These were health sciences students, second-, fourth- and sixth-year medical students, and internists. Subjects were presented with the same case of acute bacterial endocarditis as in the Patel et al. studies (Patel, Evans & Kaufman, 1990; Patel & Groen, 1986; Patel et al., 1988). In the first study, subjects were requested to study the case in order to formulate a diagnosis, to recall the case and to provide a pathophysiological explanation for the signs and

symptoms in the case. The time given for case processing was either 3 min 30 sec, 1 min 15 sec, or 30 sec.

The data on diagnostic accuracy indeed showed that experts provided better diagnoses than students under all three processing time conditions. The recall data, furthermore, showed again an intermediate effect in clinical case recall: Experts provided less elaborate recall protocols than advanced students in the long processing time condition. However, the effect disappeared in the shorter processing time conditions: Whereas the recall protocols of experts were as elaborate in all three processing time conditions, those of students and novices were less elaborate when they were given less time to study the case information. These results, thus, confirmed the expectations of Schmidt and Boshuizen that students, but not experts, need time for elaborate processing in comprehending clinical cases. Furthermore, experts were found to produce, in contrast to students, inferences summarizing two or more case elements. For example, experts used the concept sepsis, whereas students more literally recalled that the patient looked toxic, and had fever, chills, a high pulse rate and a low blood pressure. The number of these summaries was not affected by processing time. The authors interpreted these results as additional support for their hypothesis that experts, in contrast to students, process cases in encapsulated mode.

The pathophysiological explanation data also showed an intermediate effect: Fourth- and sixth-year students provided more extensive explanations than either novices, beginning students and experts. Constraining processing time resulted in less extensive explanations in students, but not in experts. The quality of the explanations was measured by comparing the subjects' explanations with a canonical explanation of the endocarditis case. This canonical explanation contained a minimal but sufficient set of biomedical and clinical knowledge which causally explained the signs and symptoms in the case. The comparison revealed that the explanations of the more experienced subjects had more concepts in common with the explanatory concepts in the canonical explanation. Thus, the experts provided highly condensed explanations of a high quality. According to the investigators, this indicated that experts explained the pathophysiology underlying the case in encapsulated form.

In the second experiment described by Schmidt and Boshuizen (1993b), a knowledge activation task preceded the reading, diagnosis and recall of the endocarditis case. In the activation task subjects were instructed to tell everything they knew about endocarditis in either 3 min 30 sec, or 30 sec. The processing time for the endocarditis case was set to 30 sec. The activation protocols showed that subjects activated considerably more prior knowledge if more time was available. For the 3 min 30 sec condition a clear expertise effect was present: The more experienced the subjects, the more knowledge they generated. Although processing time was identical for all subjects, only students were found to produce more elaborate case recall when they had

had more time to activate their knowledge. The recall data showed a clear intermediate effect in the long activation time condition, but not in the short activation time condition. Schmidt and Boshuizen considered these results as strong evidence that students, but not experts, need to process pathophysiological knowledge to comprehend a clinical case.

Conclusion

The two-world hypothesis (Patel et al., 1989a; Patel et al., 1989b; Patel & Kaufman, 1995) and the theory of knowledge encapsulation (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992, 1993a) both clearly differ from the view that biomedical knowledge has a central and integrating function in expert diagnostic reasoning (Feltovich & Barrows, 1984; Feltovich et al., 1984; Kuipers & Kassirer, 1984; Lesgold et al., 1988; Lesgold, 1984). Both of these theories rather assume that expert diagnostic reasoning relies on direct clinical association, and that biomedical knowledge is merely used to explain clinical phenomena. Therefore, both the two-world hypothesis and the theory of knowledge encapsulation can explain most of the experimental results discussed above. Through direct activation of the relevant clinical knowledge in diagnosing routine cases, experts generate accurate diagnoses in a very short time without activating more detailed biomedical knowledge. As a consequence, the case representations of experts are very condensed containing a relatively large number of high level clinical concepts. This contrasts with the case representations of students who do not immediately activate the adequate knowledge, and need to process clinical cases more elaborately. These differences in case processing by medical experts and students gave rise to the intermediate effects found in clinical case recall and pathophysiological case explanations. Experts were less elaborate in their recall and explanations, but they produced more summaries in their recall, and used more high level explanatory concepts in their explanations. In addition, constraining processing time was found to affect case representations of students but not of experts. On the other hand, it has been shown that detailed biomedical knowledge is not lost in the course of development towards medical expertise. Experts used this knowledge to explain clinical findings if they were explicitly probed to do so, or if they were confronted with complex clinical cases. The studies by Lesgold et al. (1984; 1989) on expertise in radiology, furthermore, showed that experts used more specific pathophysiological knowledge than residents while diagnosing x-ray films. However, because the x-ray films used could not be characterized as particularly difficult, this might be a consequence of the perceptual nature of the domain.

Despite their commonalities concerning the role of biomedical and clinical knowledge in medical reasoning, the two-world hypothesis and the theory of knowledge encapsulation clearly differ in their assumptions on the relations between biomedical and clinical knowledge. Whereas Patel et

al. emphasized that biomedical and clinical knowledge are two worlds apart, Schmidt and Boshuizen emphasized that biomedical and clinical knowledge are meaningfully integrated in the course of expertise development. Patel et al. only vaguely stated that the correspondences between the two worlds need to be developed in the course of medical training and practice, but they did not specify how. Schmidt and Boshuizen, by contrast, proposed a developmental theory on the restructuring of knowledge in which biomedical knowledge is assumed to be incorporated into clinical knowledge as a result of repeated use in medical problem solving. They introduced knowledge encapsulation as a learning mechanism that mediates the transition from intermediate to expert knowledge structures. As a result of these structural changes in the knowledge base, experienced physicians would be able to diagnose routine clinical cases fast and adequately.

Mechanisms of learning

The learning process of knowledge encapsulation emanates from solving clinical problems. Other authors have described learning mechanisms that are based on the performance of similar cognitive tasks. This section briefly discusses these learning mechanisms and the position of knowledge encapsulation among them.

Performance in complex task domains usually requires the completion of a sequence of sub-tasks. In problem solving tasks, these are the understanding of the problem, the planning of the steps necessary to solve the problem, and the execution of these steps. For expert problem solving each of these sub-tasks needs to be optimized. To understand a problem one needs to have the relevant domain knowledge, and to actually solve the problem one needs to apply the right procedures or rules.³ Task performance of experts is assumed to be considerably faster and more accurate than that of novices, because they are able to invoke the right knowledge at the right time, and because they have developed skilled procedures. The mechanisms involved in the development of the knowledge base and in the acquisition of skills will be successively discussed.

Problem understanding refers to the situation that a coherent problem representation is formed. Many researchers assume that experts are able to do this in a very short time because they have developed a schema-based organization of their knowledge. A schema is a knowledge structure that comprises the knowledge to interpret a class of problems or situations, as well as the knowledge to guide solution processes and behavior (Anderson, 1990; Elio & Scharf, 1990; Koedinger & Anderson, 1990; Rumelhart &

³ In the psychological literature, this distinction between knowledge of the subject matter and knowledge of procedures is known as the distinction between declarative and procedural knowledge.

Norman, 1978; Rumelhart & Norman, 1985). A clear example of the use of schemas in problem solving comes from the domain of physics (Chi, Feltovich & Glaser, 1981; Larkin, McDermott, Simon & Simon, 1980). Experts in this domain were found to be quick and accurate problem solvers because they immediately represented problems on the basis of the physics laws underlying their solution. Thus, the physics laws formed the organizing principles or schemas guiding both the classification and the solution of problems. Novices, by contrast, did not generate these schema-based representations, but represented problems rather in terms of the features that were given in the problem statement. Rumelhart and Norman (1978) proposed three general learning mechanisms involved in the transition from novice to expert knowledge structures. These are accretion, tuning and restructuring. Accretion simply refers to the accumulation of new information in memory. Tuning refers to the process of specializing, generalizing or conditionalizing schemas on the basis of the problems and situations encountered. Restructuring, finally, refers to the creation of new schemas by modifying old ones when new information does not fit into the existing schemas.

Not only does the structure of domain knowledge change in the growth towards expertise, problem solving skill is also acquired as the steps taken in problem solution are executed faster and become more efficient (e.g., Blessing & Anderson, 1996; VanLehn, 1996). The speed-up in skilled performance has found to be largely due to changes in the process leading to a solution: Experts in a task domain usually take fewer and different steps than novices to solve the problems in that particular domain. Cognitive skill acquisition, therefore, is by some authors characterized as a step-skipping process (Blessing & Anderson, 1996; Koedinger & Anderson, 1990). A typical example of the findings in cognitive skill acquisition research is the difference between experts and novices in solving linear equations: When asked to solve the equation of $2x + 3 = 9$, experts will immediately know that $x = 6/2 = 3$, whereas novices will conduct at least two more transformations to obtain this result (e.g., $2x + 3 - 3 = 9 - 3$, $2x = 9 - 3$, $2x = 6$, $2x/2 = 6/2$, $x = 6/2$, $x = 3$). Two learning mechanisms that can explain these changes in problem solving procedures are composition (Anderson, 1983, 1987) and chunking (Newell, 1990; Newell & Rosenbloom, 1981). Both mechanisms were proposed in developing a computer model of cognitive functioning in which problem solving procedures were implemented as productions. Productions are condition-action rules that produce the action if the conditions are met. Composition is the learning mechanism in ACT* (Anderson, 1983, 1987) through which a sequence of productions is collapsed into a single production that has the same effect as the sequence. Chunking, the learning mechanism in SOAR (Newell, 1990), is different from composition because new productions need not to be built from a fixed sequence of productions that lead to a solution. In SOAR, learning occurs if an impasse in problem solv-

ing is resolved: A new production is created from the results of the problem solving process and the conditions that gave rise to these results. Through repeated practice, the chunks of processing become larger and larger.⁴

In knowledge encapsulation theory, the difference between student and expert performance in medical problem solving, i.e. diagnosing a clinical case, was characterized as elaborate processing of detailed biomedical knowledge versus immediate activation of relevant clinical knowledge (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992, 1993b). Medical students were found to actively reason with the knowledge acquired in the course of their training to verify diagnostic hypotheses. In this process, general problem solving procedures, such as inferencing, are used to make sense out of the data. Expert physicians, on the other hand, did not engage in this kind of elaborate processing, but immediately activated and verified the relevant set of diagnostic hypotheses in proceeding through a case. Their performance, therefore, could be considered as a step-skipping process. However, in contrast to the learning mechanisms of composition (Anderson, 1983, 1987) and chunking (Newell, 1990; Newell & Rosenbloom, 1981), the learning mechanism of knowledge encapsulation as proposed by Schmidt and Boshuizen (1992) does not change the procedures that are used in case processing. Knowledge encapsulation changes the structure of the domain knowledge so that case comprehension can proceed at a faster rate.

Overview of the present studies

In Chapters 2 through 5, four experiments will be presented that have been performed to test the notion of knowledge encapsulation. They partly replicate and extend experiments done by Schmidt and Boshuizen (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1993b). The results of clinical case representation studies, and especially the intermediate effect, played a crucial role in supporting the idea of knowledge encapsulation (Schmidt & Boshuizen, 1992; Schmidt & Boshuizen, 1993b). Schmidt and Boshuizen assumed that expert physicians represent a clinical case in a more condensed form than students of an intermediate level of expertise, because they do not need elaborate processing of biomedical knowledge to comprehend a clinical case, but process the data at a higher level of aggregation. The manipulations of processing time in the 1993b study of Schmidt and Boshuizen corroborated these assumptions showing that under shorter processing times the representations of students were affected, but not those of experts. However, this study was performed with only one clinical case. To test the robustness of the intermediate effect and related phenomena in case repre-

⁴ Note that chunking in SOAR is different from chunking described in the general psychological literature (e.g., Chase & Simon, 1973; Miller, 1994), because it acts upon problem solving procedures and not on data being interpreted.

sentation we performed two clinical case recall studies and a pathophysiological explanation study using different materials, additional experimental conditions, and more detailed methods of analysis.

In Chapter 2, a study is presented in which we investigated the robustness of the intermediate effect in clinical case recall by manipulating the recall instructions given to the subjects (i.e. incidental vs. intentional), as well as the format of the lab data in the case descriptions (numerical vs. interpreted). Under these conditions, subjects of three levels of expertise studied, diagnosed and recalled three clinical cases. Processing time was recorded.

The other recall study we performed was a replication of the 1993b study of Schmidt and Boshuizen in which the time available for case processing was manipulated. However, instead of the endocarditis case used by Schmidt and Boshuizen, we used four cases from different specialty areas in internal medicine. Subjects of four levels of expertise studied the cases under three different time constraints, and provided their diagnoses and recall. This study is presented in Chapter 5.

In Chapter 3, a pathophysiological explanation study is presented that aimed to investigate in detail the structure of the causal knowledge underlying clinical case representations in students and physicians. In this study four clinical cases were presented and processing time was manipulated. The method used for analyzing the explanation protocols was more precise than in earlier studies (Boshuizen, 1994; Schmidt & Boshuizen, 1993b) and considered in detail the concepts used, as well as the links between these concepts.

In Chapter 4, another perspective was taken on the study of knowledge encapsulation: We did not focus on the knowledge used in the representation of clinical cases, but on the content and structure of the knowledge that is subsumed under encapsulating concepts. Schmidt and Boshuizen (1992; 1993b) assumed that as a consequence of experience in medical practice, coherent networks of biomedical knowledge relating causes and consequences of disease, would become encapsulated into more comprehensive clinical concepts. They also assumed that experts would be able to unfold the encapsulated knowledge whenever required for task performance. Therefore, we expected that if students and physicians are asked to explain current clinical concepts that could be characterized as encapsulating concepts, the physicians would provide more elaborate, qualitatively better and more fluent explanations. In addition, we expected that clerks who just finished a period of relevant clinical experience would do better than advanced students who had not yet had clinical experience.

In Chapter 6, finally, the idea of knowledge encapsulation is revisited in the context of the results of the present studies. This results in a reformulation of Schmidt and Boshuizen's theory on the development of medical expertise (1993a) by providing a more precise account of the changes in the

knowledge structures in the growth from student to expert physician. Implications for future research and medical education are discussed.

References

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, 94(2), 192-210.
- Anderson, J. R. (1990). *Cognitive psychology and its implications* (third edition ed.). New York: W. H. Freeman and Company.
- Balla, J. I., Biggs, J. B., Gibson, M., & Chang, A. M. (1990). The application of basic science concepts to clinical problem-solving. *Medical Education*, 24, 137-147.
- Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-Based Learning: An approach to medical education*. New York: Springer Publishing Company.
- Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 576-598.
- Bordley, J., & Harvey, A. M. (1976). *Two centuries of American Medicine*. Philadelphia, PA: W. B. Sanders Company.
- Boshuizen, H. P. A. (1989). *De ontwikkeling van medische expertise: een cognitief-psychologische benadering* [The development of medical expertise; a cognitive psychological approach]. Doctoral dissertation. Haarlem: Thesis publishers.
- Boshuizen, H. P. A. (1994). Cognitive effects of practical experience in high- and low-achieving medical students. *Learning and Instruction*, 4, 313-329.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Bouman, L. N., & Snellen-Balendong, H. A. M. (1996). Basisvakken: een basis waarvoor? [Basic sciences: The basis for what?]. *Nederlands Tijdschrift voor de Geneeskunde*, 140(21), 1134-1137.
- Cavazos, L. F. (1984). Basic science studies: Their purpose in medical education. *Journal of Medical Education*, 59, 763-764.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Dawson-Saunders, B., Feltovich, P. J., Coulson, R. L., & Steward, D. E. (1990). A survey of medical school teachers to identify basic biomedical concepts medical students should understand. *Academic medicine*, 65(7), 448-454.
- Duffy, J. (1993). *From humors to medical science: A history of American medicine*. Chicago: University of Illinois Press.
- Elio, R., & Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science*, 14, 579-639.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving*. Cambridge, MA: Harvard University Press.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, Massachusetts: MIT Press.
- Feltovich, P. F., Spiro, R. J., & Coulson, R. L. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling*, (pp. 113-172). Cambridge, MA: The MIT Press.
- Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt & M. L. De Volder (Eds.), *Tutorials in problem-based learning. New directions in training for the health professions*, (pp. 128-142). Assen/Maastricht: Van Gorcum.
- Feltovich, P. J., Coulson, R. L., Spiro, R. J., & Dawson-Saunders, B. K. (1992). Knowledge application and transfer for complex tasks in ill-structured domains: Implications for in-

- struction and testing in biomedicine. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 213-244). Berlin: Springer Verlag.
- Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence: the first decade*, (pp. 275-319). Reading, MA: Addison-Wesley Publishing Company.
- Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1993). Learning, teaching and testing for complex conceptual understanding. In N. Frederiksen, R. Mislevy, & I. Bejar (Eds.), *Test theory for a new generation of tests*, (pp. 180-217). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Flexner, A. (1910). *Medical Education in the United States and Canada* (Report No. 4). New York: The Carnegie Foundation for the Advancement of Teaching.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14, 511-550.
- Kuipers, B., & Kassirer, J. P., (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science*, 8, 363-385.
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Lesgold, A., Robinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill; Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*, (pp. 311-342). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory 'Essays in honor of Gordon Bower'*, (pp. 31-60). San Francisco, New York: W. H. Freeman and Company.
- Maddison, D. (1978). What's wrong with medical education? *Medical Education*, 12(2), 97-106.
- Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 101(2), 343-352.
- Muzzin, L. J., Norman, G. R., Jacoby, L. L., Feightner, J. W., Tugwell, P., & Guyatt, G. H. (1982). Manifestations of expertise in recall of clinical protocols. *Proceedings of the 21st Conference on Research in Medical Education*, 163-168.
- Neame, R. L. B. (1984). The preclinical course of study: Help or hindrance? *Journal of Medical Education*, 59, 699-707.
- Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Medical Education*, 15, 315-322.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*, (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Norman, G. R., Jacoby, L. L., Feightner, J. W., & Campbell, E. J. M. (1979). Clinical experience and the structure of memory. *Proceedings of the 18th Conference on Research in Medical Education*, 214-219.
- Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125, 1063-1068.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989a). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling*, (pp. 53-112). Cambridge, MA: The MIT Press.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989b). Reconciling basic science and clinical reasoning. *Teaching and Learning in Medicine*, 1(3), 116-121.
- Patel, V. L., Evans, D. A., & Kaufman, D. R. (1990). Reasoning strategies and the use of biomedical knowledge by medical students. *Medical Education*, 24, 129-136.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V. L., & Groen, G. J. (1991a). Developmental accounts of the transition from medical student to doctor: some problems and suggestions. *Medical education*, 25, 527-535.
- Patel, V. L., & Groen, G. J. (1991b). The general and specific nature of medical expertise: A critical look. In A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*, (pp. 93-125). Cambridge: Cambridge University Press.

- Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition*, 18(4), 394-406.
- Patel, V. L., Groen, G. J., & Scott, H. M. (1988). Biomedical knowledge in explanations of clinical problems by medical students. *Medical Education*, 22, 398-406.
- Patel, V. L., & Kaufman, D. R. (1995). Clinical reasoning and biomedical knowledge: Implications for teaching. In J. Higss & M. Jones (Eds.), *Clinical reasoning in the health professions*, (pp. 117-128). Oxford, UK: Butterworth Heinemann.
- Patel, V. L., & Medley-Mark, V. (1986). Relationship between representation of textual information and underlying problem representation in medicine. *Cognitive research series*. Montreal: McGill University, Centre for Medical Education. (CME86-CS1)
- Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: Three modes of learning. In R. Klatzky, & J. W. Cotton (Eds.), *Semantic factors in cognition*, (pp. 37-60). Hillsdale, NJ: Erlbaum.
- Rumelhart, D. E., & Norman, D. A. (1985). Representation of knowledge. In A. M. Aitkenhead & J. M. Slack (Eds.), *Issues in cognitive modeling*, (pp. 15-62). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-282). New York, NY: Springer Verlag.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), 205-221.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, 21(3), 338-351.
- Schmidt, H. G., Boshuizen, H. P. A., & Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 139-145.
- Schmidt, H. G., & De Volder, M. L. (1984). *Tutorials in problem-based learning. New directions in training for the health professions*. Assen/Maastricht: Van Gorcum.
- Simpson, M. (1972). *Medical education: A critical approach*. London: Butterworths.
- VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513-539.

Robustness of the intermediate effect in clinical case recall¹

In the present study, the influence of recall instruction and lab data format on free recall of clinical cases was investigated in medical students and expert physicians. Subjects studied, diagnosed, and recalled three clinical cases. No differences in task performance were found between intentional and incidental recall instructions, nor between cases with numerical and interpreted lab data. The results showed that in all three cases intermediate students provided more elaborate recall protocols than both beginning students and experts. Students, furthermore, studied the cases for a longer time, diagnosed the cases less accurately, and produced fewer summaries in their recall protocols than experts. These findings support the notion that experts process clinical cases largely automatically using knowledge in encapsulated form, whereas students process clinical cases more elaborately applying detailed pathophysiological knowledge.

Theories of text-processing generally assume that prior knowledge affects new information encoding and retrieval (e.g., Bransford & Johnson, 1972; Graesser & Clark, 1985; Voss & Bisanz, 1985; Kintsch, 1988). New information from text and prior knowledge interact to construct a mental representation of the text. Studies using free recall of meaningful texts assume that the recall reflects content and structure of that mental representation. Since experts have access to more elaborate and more coherent knowledge structures, they are conjectured to recall more information from the text than novices. Research on expertise in domains such as chess (de Groot, 1965; Chase & Simon, 1973), baseball (Spilich, Vesonder, Chiesi & Voss, 1979), bridge (Charness, 1979), electronic engineering (Egan & Schwartz, 1979), and

¹ Parts of this chapter appeared as Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (1994). *The influence of task instruction and lab data format on clinical case recall*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, April 4-8. (ERIC Document Reproduction Service No. ED 367710)

computer-programming (McKeithen, Reitman, Rueter & Hirtle, 1981) supports this hypothesis.

Studies on expertise in the domain of medicine, however, do not produce such a clear-cut linear relationship. In these studies, subjects of different levels of expertise are presented with a clinical case, usually consisting of a description of a patient's history, the complaints and symptoms, the results of the physical examination, the relevant laboratory data and the results of additional examinations. The subjects are required to study the case and to recall whatever they remember. Only a few of these studies reported a positive relationship of recall with level of expertise (e.g., Coughlin & Patel, 1987; Hobus, Schmidt, Boshuizen & Patel, 1987; Norman, Brooks & Allen, 1989). Many studies, however, failed to find the expected increase in recall with higher expertise level (e.g., Hassebrock, Johnson, Bullemer, Fox & Moller, 1993; Muzzin, Norman, Feightner & Tugwell, 1983; Patel & Medley-Mark, 1986). These findings were difficult to interpret in the context of what is known about the relationship between the amount of prior knowledge and recall. One would expect that clinical experience would be synonymous with more extensive and more coherent knowledge of the medical domain, and hence with more elaborate case recall. Some researchers observed that more experienced subjects used less time to study or recall case information, and reported parameters such as recall per unit study time or recall per unit recall time, which *did* have a positive, linear relationship with level of expertise (e.g., Boshuizen, 1989; Claessen & Boshuizen, 1985; Muzzin, Norman, Jacoby, Feightner, Tugwell & Guyatt, 1982; Muzzin et al., 1983; Norman, Jacoby, Feightner & Campbell, 1979). Because theories of text processing predict a monotonic relation between recall and expertise, it was even more striking that several studies demonstrated an inverted U-shaped relation between level of expertise and amount of clinical case recall (e.g., Boshuizen, 1989; Patel & Frederiksen, 1984; Patel & Groen, 1991; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993b). The finding that subjects of an intermediate level of expertise provide more elaborate recall protocols than expert physicians is called the *intermediate effect* in clinical case representation studies (Schmidt, Boshuizen & Hobus, 1988).

In an attempt to summarize the results of the recall studies in the medical domain, Boshuizen (1989) performed a meta-analysis on outcomes of 11 recall studies conducted between 1979 and 1987, which suggested that the inverted U-shaped relation indeed better describes the combined results than a monotonically increasing function. This led to the conclusion that the intermediate effect is a finding of potential theoretical significance, requiring further experimentation and possibly theoretical adjustment.

Schmidt and Boshuizen (1993b), then, hypothesized that the decrease in recall performance beyond the intermediate level might result from different ways of case processing by subjects of different levels of expertise due to knowledge restructuring in associative memory. In diagnosing a case, in-

intermediate-level students reason through their rich, elaborated causal networks explaining the causes and consequences of disease in terms of general underlying biological or pathophysiological processes. As a result, these students will construct an elaborate representation of the clinical case at hand. This processing is considered to be non-automatic, involving time-consuming search. However, through extensive and repeated application of acquired knowledge, and particularly through exposure to patient problems, networks of highly related lower order concepts become encapsulated into a smaller number of higher level concepts or diagnostic labels.

The cognitive mechanism responsible for this kind of restructuring of the knowledge base is referred to as *knowledge encapsulation*. Knowledge encapsulation is defined as the subsumption or "packaging" of lower-level, detailed concepts and their interrelations in an associative net, under a smaller number of higher-level concepts with the same explanatory power (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992; Schmidt & Boshuizen, 1993b). Thus, assuming that experts' knowledge is organized in an encapsulated way, experts' case representations may contain encapsulating concepts rather than the more detailed concepts activated by students of an intermediate level of expertise. Because the concepts used by students to comprehend information in a case are more fine-grained, their representation of the case, and, hence, their recall, is more fine-grained.

Empirical evidence for this assumption has been found in a study in which case processing in lay persons, medical students and experts was experimentally manipulated (Schmidt & Boshuizen, 1993b). In this study, subjects of five different levels of expertise were requested to study a clinical case of bacterial endocarditis for a fixed time in order to produce a diagnosis, and, in addition, to write down their recall and pathophysiological explanation for the signs and symptoms in the case. Depending on the particular experimental condition, subjects were given the opportunity to study the case for 3 min. 30 sec, 1 min. 15 sec, or 30 sec. It was found that medical experts indeed provided less elaborate recall protocols than advanced students when given 3 min. 30 sec to process a case. However, the experts' recall performance was not affected by the sizable reduction in processing time in the two other conditions, whereas the students' performance was. Thus, the intermediate effect was observed when subjects had ample time to study the case and disappeared when processing time was constrained. These results confirmed the hypothesis that by reducing processing time, students are restricted in elaborating on their biomedical knowledge and, hence, construct less extensive case representations. By contrast, the experts seem to process the clinical case in an encapsulated mode and build a representation fast, independent of the particular processing time condition. In addition, Schmidt and Boshuizen found more *summaries* in the experts' case recall (a summary being an inference based on two or more propositions from the

text of a case), which they saw as further evidence for the idea that experts process a case in encapsulated mode.

The intermediate effect and related phenomena such as the number of summaries in case recall thus play a crucial role supporting the idea of knowledge encapsulation as a developmental construct and, as a consequence, Schmidt and Boshuizen's (1993a) theory on the development of expertise in medicine. If the intermediate effect is to play such an important role, it may be useful to consider the other studies in this domain in more detail, if only to find clues as to why some of these fail to disclose such effect.

A closer look at the free recall studies conducted in medicine (see Table 2.1) leads to the conclusion that there are considerable methodological differences among them. Levels of expertise of subject groups, medical texts presented, tasks, instructions and analyses vary greatly. This is most clearly the case with regard to the subjects participating in the various experiments. Some studies describe second-year medical students as novices (Boshuizen, 1989; Claessen & Boshuizen, 1985; Coughlin & Patel, 1987; Muzzin et al., 1982; Norman, et al., 1979), whereas in others laymen (Norman et al., 1989; Patel & Groen, 1991; Schmidt & Boshuizen, 1993b), beginning medical students (Patel & Medley-Mark, 1986) or even recently graduated M.D.'s (Hobus et al., 1987) are considered novices. Thus, whether subjects are considered novices, intermediates or experts, is defined relative to the other subject groups participating in an experiment. Therefore, the different recall studies can only be interpreted meaningfully if the nature of the subject groups is explicitly taken into account. This means that an absolute rather than a relative definition of expertise levels should be applied to compare the different studies. According to Schmidt and Boshuizen (1992), subjects who are about to enter clerkships² can be considered real intermediates and can, therefore, be expected to display the highest recall. Thus, the intermediate effect can only be observed, if studies include at least three types of subjects: Beginning clerks considered to be subjects of an intermediate level of expertise, subjects with considerably less medical training, and subjects with considerably more medical training than these beginning clerks. These three groups of subjects were *not* used in studies of Claessen and Boshuizen (1985), Coughlin and Patel, (1987), Hassebrock et al. (1993), Hobus et al. (1987), Muzzin et al. (1982, 1983), Norman et al. (1979, 1989), and Patel and Groen (1991) and, thus, could explain why in these studies intermediate effects were not observed. Obviously, if only two expertise levels are distinguished in an experiment, relationships other than linear are by definition impossible and, hence, nothing can be said about the presence or absence of an intermediate effect. This was the case in some studies that reported an increasing recall perfor-

² Medical curricula in North America consist of a four-year program. Students in these programs enter their clerkships in the second or third year. In the Netherlands, by contrast, medical curricula require six years of training and students enter their clerkships in the fifth year.

mance with level of expertise (Coughlin & Patel, 1987; Hobus et al., 1987; Norman et al., 1989; Patel et al., 1986).

Another important factor concerns the medical cases presented in the different studies. In most recall studies subjects were presented clinical cases in which a patient's history, complaint, physical examination, relevant lab data and additional findings were described. However, Norman et al. (1989) presented their subjects exclusively with records of lab data, Patel & Frederiksen (1984) presented two basic science textbook excerpts, and Hobus et al. (1987) presented slides with a patient's portrait, chart and complaint. As will be discussed subsequently, the use of these different materials may be responsible for the fact that experts provided more elaborate recall protocols than advanced students in two of these studies (Hobus et al., 1987; Norman et al., 1989).

Furthermore, the experimental tasks and the order in which these tasks are performed are important aspects in which the clinical case recall studies differ from each other. Coughlin and Patel (1987), Muzzin et al. (1982, 1983) and Norman et al. (1979), for instance, required subjects to study a case, to recall the text as accurately as possible and then to provide a diagnosis. Patel and Medley-Mark (1986) and Schmidt and Boshuizen (1993b), on the other hand, asked subjects to diagnose the case first before requiring recall. Boshuizen (1989) and Hassebrock et al. (1993) instructed subjects to think aloud with the pertinent task to produce an accurate diagnosis and then to recall the case. Finally, Claessen and Boshuizen (1985) instructed subjects to think aloud to formulate a diagnosis, but, in addition, to categorize case information as important or unimportant, before requiring recall. Thus, in some studies emphasis was on the recall task, whereas in other studies the diagnostic task was emphasized and recall was more or less incidental. Because in medical practice clinical data are usually processed to diagnose or monitor the condition of a patient, the recall of a written clinical case most clearly reflects expertise if subjects represent that case as the result of a diagnostic process. Active memorization of clinical cases, by contrast, does not reflect the way experts usually process clinical data and, as a consequence, could actually mask the effect of expertise (Norman et al., 1989).

Finally, the recall studies also differed in the time subjects were given to study a clinical case. Available study time varied from 30 seconds in some studies (Hobus et al., 1987; Schmidt & Boshuizen, 1993b) to essentially unrestricted processing time conditions in others (e.g., Claessen & Boshuizen, 1985; Hassebrock et al., 1993; Norman et al., 1989). Since time available to perform a particular task may be expected to influence how well that task is performed, differences in processing time between the different studies may have been responsible for at least some of the inconsistencies in the findings.

If only those studies are considered in which subjects of at least three levels of expertise participated, including an intermediate group of students

Table 2.1. Summary of recall studies in the domain of medical expertise.

Study	Subjects	Materials	Tasks	Study time	Effect of expertise level on recall
Norman, Jacoby, Feightner & Campbell (1979)	5 2nd-year medical students 5 1st-year residents in family medicine 5 3rd-year residents in family medicine 5 family physicians	4 clinical cases: gallbladder disease, lung cancer, and 2 cases not possible to diagnose accurately	study, explicit intentional recall, diagnosis with likelihood ratings	free 65-210 sec	no significant effect • recall per unit study time: positively linear effect
Muzzin, Norman, Jacoby, Feightner, Tugwell & Guyatt (1982)	8 2nd-year med students 8 internists	8 cardiorespiratory cases: 4 typical, 4 atypical	listen to and read case, explicit intentional recall, diagnosis	fixed: 2 min	no significant effect • recall per unit recall time: experts > students
Patel & Frederiksen (1984)	8 1st-year medical students 8 2nd-year medical students 8 internists	2 medical text book excerpts on fever and on cancer	write 1 page about subject, study text, explicit intentional recall	fixed	inverted U-shaped effect
Patel, Groen & Frederiksen (1986) Reanalysis of Muzzin et al. (1982)	6 2nd-year medical students 6 internists	4 cardiorespiratory cases: 2 typical, 2 atypical	listen to and read case, explicit intentional recall, diagnosis	fixed: 2 min	• typical cases: students > experts • atypical cases: no significant effect
Claessen & Boshuizen (1985)	7 2nd-year medical students 10 3rd-year medical students and two groups of 8/9 4th-year medical students and 6/7 5th-year medical students	4 clinical cases: 2 typical, 2 atypical of pancreatitis and prostatitis; case information was presented on cards	read cards, thinking aloud for diagnosis and sorting cards into important and unimportant, recall	free negatively linear relation with expertise level	no significant effect • recall per unit study time: positively linear effect
Patel & Medley-Mark (1986)	9 beginning medical students 7 2nd-year medical students 10 4th-year medical students 7 internists	2 clinical cases: stomach carcinoma and endocarditis	read, diagnosis, summarize, answer some questions	fixed: 2 min 30 sec	inverted U-shaped effect
Coughlin & Patel (1987)	16 2nd-year medical students 16 family physicians	4 clinical cases: 2 in typical structure, 2 in random structure of endocarditis and arthritis	read, explicit intentional recall, diagnosis	fixed: 3 min 30 sec	experts > students

Hobus, Schmidt, Boshuizen & Patel (1987)	17 almost/just graduated medical students 18 family physicians	32 clinical cases: each case presented on 3 slides with respectively a patient's portrait, chart and complaint	after exposure to slides, diagnosis; after 16 cases incidental recall of these cases; after 32 cases recall of last 16 cases	fixed exposure time: mean of 32 sec	experts > students
Boshuizen (1989) 1st experiment	16 2nd-year medical students 16 4th-year medical students 16 5th-year medical students 16 family physicians	4 clinical cases: 2 typical, 2 atypical of pancreatitis and prostatitis; case information was presented on cards	read cards, thinking aloud for diagnosis, recall	free: ±650 - ±900 sec 2nd & 4th year medical students > 5th year medical students & physicians	inverted U-shaped effect • recall per unit study time: positively linear effect
Norman, Brooks & Allen (1989), 1st experiment	5 research assistants in psychology 6 2nd-year medical students 5 specialists	16 sets of laboratory data: 4 patient data, 4 random data presented in standard or scrambled format	read, explicit intentional recall	free: ±70 - ±240 sec inverted U-shaped relation with expertise level	• patient data: positively linear effect
Norman, Brooks & Allen (1989), 2nd experiment	6 3rd-year medical students 6 specialists	6 sets of real patient laboratory data	1. read, diagnose 2. read, diagnose, recall 3. read, memorize, recall	free no significant effect	• diagnose condition: experts > students • memorize condition: students ≥ experts
Patel & Groen (1991) Experiment 1 (pp. 110)	subjects differing in premedical training: 1st-year college students premedical-year students degree students	1 clinical case of endocarditis	read, explicit intentional recall, diagnosis	?	inverted U-shaped effect
Hassebrock, Johnson, Bullemer, Fox & Moller (1993)	4 3rd-year medical students 4 trainees pediatric cardiology 4 pediatric cardiologists	2 clinical cases of congenital heart disease; case information was presented in a matrix	selection of information, thinking aloud for diagnosis, immediate recall, recall after one week	free	• immediate recall: no significant effect • recall after one week: negatively linear effect
Schmidt & Boshuizen (1993b)	24 1st-year health sciences students 24 2nd-year medical students 24 4th-year medical students 24 6th-year medical students 24 internists	1 clinical case of endocarditis	study, diagnosis, recall, pathophysiological explanation	fixed: 3 min 30 sec, 1 min 15 sec or 30 sec	inverted U-shaped effect • 3 min 30 sec: inverted U-shaped effect with linear component • 1 min 15 sec: no significant effect • 30 sec: positively linear effect

Note: • indicates the effect of expertise on recall for a specified condition.

who are about to enter the clinical clerkships, four out of five studies showed an intermediate effect in recall of medical information (Boshuizen, 1989; Patel & Frederiksen, 1984; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993b). This suggests that the intermediate effect in clinical case recall is indeed a stable effect. The exception, however, was a study performed by Norman et al. (1989), showing a positive linear recall function with expertise. Interestingly, not only were the results obtained in this study different, so too were the experimental materials and tasks that were applied. Whereas in most of the other studies subjects were instructed to diagnose and recall clinical cases (Boshuizen, 1989; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993b), in the study of Norman et al. subjects were instructed to memorize and recall protocols of numerical laboratory data. Therefore, it seems useful to describe the experiments performed by Norman et al. in more detail.

In the first experiment Norman et al. (1989) presented eight protocols of laboratory data to medical specialists, intermediate-level second-year medical students and psychology students. Each protocol contained the numerical values of 20 lab data. The values were either based on realistic patient data or on random values. Subjects were instructed to read a protocol and to recall the lab data as accurately as possible. The time allowed to read the data was free. The results showed that the number of lab data recalled increased with level of expertise if the lab data were based on realistic patient data. However, if random values were assigned to the lab data, no difference in amount of recall was found for the three groups of expertise; experts and medical students performed as poorly as novices. In addition, experts and medical students used more time to study the random values data than the realistic patient data. However, the effects were not as large as expected. Norman et al. originally assumed that experts in contrast to students would automatically recognize the disease patterns in the realistic patient data and, hence, would recall these data much better than students. This expectation was based on the finding that chess experts process and recall familiar patterns of chess pieces largely automatically (e.g., Chase & Simon, 1973). However, although the experts recalled more lab data than the medical students in the realistic patient data condition, a separate analysis for these two groups revealed that the difference was not significant. Norman et al. suggested that the perception of diagnostic patterns in visual arrays of lab data was not so automatic and effortless as expected, and hypothesized that the effect of expertise on recall might be stronger when subjects were directly instructed to engage in a diagnostic task rather than being instructed to memorize the data. This hypothesis was tested in a second experiment.

In this second experiment, third-year students and medical specialists were subsequently presented six protocols of lab data with three different types of task instructions. For the first two protocols, subjects were requested to read the protocols, to describe the patient's problem, and to relate the in-

dividual values to this problem. The experiment started with this condition to emphasize that the investigators were primarily interested in the way subjects formulated the problem. The instructions for the second two protocols were similar. However, after the discussion of each protocol subjects were asked whether there were any specific lab values which they could recall. This second condition was the diagnose condition with incidental recall. For the final two protocols, subjects were instructed to ignore the patient's problem when reading the lab data, and instead to memorize the specific values of the data and to recall these as accurately as possible. This condition was the memorize condition with intentional recall. The reading time was free in all three conditions. It was found that the experts recalled more lab data in the diagnose condition than in the memorize condition; conversely, the third-year students recalled twice as many data in the memorize condition than in the diagnose condition. In the memorize condition the recall performance of experts and students was about equal. In the diagnose condition, however, the recall performance of experts was highly superior. Based on these findings, the authors concluded that an intentional recall task with memorize instructions is not a good method to probe the way information is usually processed in medical practice. Especially difficult clinical data such as laboratory protocols will only be superficially interpreted if memorized rather than being processed for a diagnosis.

Among case recall studies of medical expertise, the findings of Norman et al. (1989) are clearly the exception. The question, then, is why this is so. Norman and colleagues themselves suggest that the diagnostic interpretation of lab data is generally so difficult that it requires an effortful analysis even for expert physicians. Because lab data, in their view, cannot be processed in an automatic, pattern-recognition-like fashion, experts pay attention to the data in an analytical way. This kind of processing would lead to a more elaborate representation of the case and hence, to better recall. This explanation is in agreement with the Schmidt-Boshuizen notions of medical expertise development. These authors also assume that, if the diagnostic task is difficult and cannot be solved by pattern-recognition, experts engage in effortful and elaborate analysis of a case, much like intermediate-level students do, resulting in more detailed recall (Schmidt and Boshuizen, 1993a).

However, the interpretation of numerical lab data in the study of Norman et al. (1989) might be unrealistically difficult: Norman et al. presented their subjects with a standard array of laboratory data, without any further information about the patient's condition. This is, however, not the way in which lab data are used in daily practice. Usually, physicians request lab data to test, or rule out, diagnostic hypotheses coming up during history taking and physical examination. Furthermore, lab tests are regularly conducted to monitor a patient's condition and to evaluate the effects of treat-

ment. In these situations, physicians thus review lab protocols in the light of explicit diagnostic hypotheses. Only rarely, in extreme cases, do physicians rely on lab data alone to solve a diagnostic problem. In the Norman et al.'s 1989 study, the subjects did not even know whether the patient was a man or a woman, while at least for some lab data the normal ranges of values are different for the two sexes.³

Our conjecture, therefore, is that experts process lab data in pattern-recognition mode, if these data are put in the natural context of a patient's history and physical examination. This kind of information processing, in turn, would lead to more condensed recall of such a case as compared with intermediate level students. In addition, we assumed that under these processing conditions it would not matter whether subjects are required to diagnose a case or to memorize it. Because processing is automatic for experts, the resulting representation would be the same under both instructions. If this is true, the recall patterns under both conditions must be similar, as opposed to the Norman et al. (1989) findings.

To test these hypotheses, subjects of three levels of expertise were presented with three cases describing a patient's history, physical examination, and lab data. They were either requested to process the case for diagnosis or for recall. Diagnostic judgments, recall and processing times were recorded.

Crucial to Norman et al.'s point of view is the expectation that experts produce more elaborate recall because they have to actively interpret the lab data in a non-automatic fashion. This would imply that, if experts had to process lab data *that were already interpreted for them*, automatic processing would be possible, resulting in more condensed recall. To test this expectation, cases were included containing the numerical lab data in interpreted format. Two examples: An erythrocyte sinking rate (ESR) of 2 mm/h in the numerical lab data condition was described as a "low normal ESR" in the interpreted lab data condition; the values of the five lab tests that constituted the analysis of blood gasses (i.e. pH, PO₂, pCO₂, HCO₃⁻-concentration and O₂-saturation) were interpreted as "analysis of blood gasses shows metabolic compensation for respiratory alkalosis". An example of a case description in both numerical and interpreted lab data format is provided in Appendix A. Norman et al.'s theory would predict different recall patterns for both types of cases, due to different processing conditions; our assumption is that both versions of a case are processed automatically and differences in recall are, therefore, not to be expected.

³ This applies, for example, to the serum creatinine test that was presented in the lab protocols of Norman et al. (1989).

Method

Subjects

The two student groups participating in the experiment consisted of 16 second-year, and 16 fourth-year medical students of Maastricht University, the Netherlands. The expert group consisted of 16 internists from five different hospitals with at least four years of experience in internal medicine, and, on average, 15 years of experience. Subjects received a small compensation for their participation.

Materials

The materials consisted of three clinical case descriptions and two blank response sheets after each case. Each clinical case description reported some contextual information, the main complaint, findings from history taking and physical examination, the relevant laboratory data and some additional findings. The three cases, a case of stomach carcinoma, a case of pheochromocytoma and a case of heart failure, were based on actual patients and were presented in a standard medical format. Except for pheochromocytoma, these cases are fairly common in medical practice. In their original format the cases contained a fairly large number of numerical lab data relative to the other case information. The pheochromocytoma case and the heart failure case were also rewritten so that they contained lab data in an interpreted format. The case descriptions were about half a page in length and consisted of 42, 33, 43 propositions for the three standard cases respectively, and 32 and 40 propositions for the rewritten pheochromocytoma and heart failure cases. Most of the case propositions were directly relevant to diagnosis. The descriptions of the case of heart failure in numerical and interpreted format are provided in Appendix 2.1.

Procedure

Subjects were instructed to study a case for up to 3 minutes to formulate a diagnosis. Before the presentation of the first clinical case (the case of stomach carcinoma) half of the subjects was told that they would subsequently be asked to write down whatever they remembered from the case. This was the *intentional recall condition*.⁴ The other half of the subjects was not aware

⁴ In the present study we slightly modified the intentional recall instructions given to the subjects as compared to the instructions given in the memorize condition in the study of Norman et al. (1989). Norman et al. instructed their subjects to ignore the patient's problem when reading the lab data. However, when subjects are presented with clinical cases including a patient's complaints, symptoms and signs, as was the case in the present experiment, it is very unlikely that experts would be able to ignore the patient's problem. All subjects participating in the present study, therefore, knew that the study involved the diagnoses of clinical cases. In the intentional recall condition, however, the recall task was emphasized.

that they would be asked to recall the case. This was the *incidental recall condition*.

For the next two cases all subjects received similar instructions as given to the subjects in the intentional recall condition, but now with emphasis on the diagnostic task. The second case the subjects were asked to diagnose and recall was the case of pheochromocytoma. Half of the subjects were presented the case in numerical lab data format, and the other half had to study it in interpreted lab data format. Finally, the subjects were asked to diagnose and recall the case of heart failure. Again, half of the subjects were presented the case in numerical lab data format, and the other half in interpreted lab data format. Assignment of subjects to conditions was balanced. An overview of the experimental design is provided in Table 2.2.

The actual study times were recorded. Subjects were free to use as much time as they needed to write down their diagnosis and recall.

Analysis

Diagnoses were scored on a scale ranging from 0 (completely inaccurate diagnosis) to 6 (completely accurate diagnosis). For each case points were awarded for accurate diagnostic elements which summed up to 6. For example, the complete diagnosis for the heart failure case was "left- and right-sided heart failure based on mitral valve insufficiency and atrial fibrillation". Two points were given when the diagnosis contained the term "heart failure". One additional point was given for each of the following elements: "left-sided", "right-sided", "mitral valve insufficiency", and "atrial fibrillation".

Based on a technique of proposition analysis for medical protocols (Patel & Groen, 1986), recall protocols were segmented into small, meaningful information units or propositions. These propositions consisted of two concepts connected by a qualifier, such as, *specification (spec)*, *location (loc)* or *time information (temp)*. For instance, the protocol fragment "(...) The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla. (...)" consists of five propositions: 1. heart-spec (enlarged to all sides); 2. auscultation-spec (murmur); 3. murmur-spec (holosystolic); 4. murmur-loc (at the apex); 5. at the apex-spec (radiating towards the axilla). For each proposition in the free recall, it was decided whether it matched any proposition, or combination of propositions, in the stimulus material. The total recall score was determined by counting the number of literally recalled propositions, paraphrases, slightly imprecisely recalled propositions, and inferences matching one or more propositions in the case. The number of summaries, i.e. inferences referring to more than one proposition in the case, was counted separately as well. For example, the five propositions in the above recall fragment matched literally with the original text giving a recall score of 5, and a summary score of 0. However, the same information might be recalled as "The heart is en-

Table 2.2. Experimental Design.

Case	Subjects*	Task instructions	Lab data format
1. Stomach carcinoma	1 - 8	Intentional	Numerical
	9 - 16	Incidental	Numerical
2. Pheochromocytoma	1 - 4 and 9 - 12	Intentional	Numerical
	5 - 8 and 13 - 16	Intentional	Interpreted
4. Heart failure	1 - 4 and 9 - 12	Intentional	Interpreted
	5 - 8 and 13 - 16	Intentional	Numerical

* Assignment of subjects to experimental conditions for each expertise group.

larged to all sides and auscultation reveals mitral valve insufficiency." In this latter recall fragment, the proposition "auscultation reveals mitral valve insufficiency" summarizes four propositions from the original text (auscultation reveals a holosystolic murmur at the apex radiating towards the axilla) giving a recall score of 2 and a summary score of 1.

The protocols were scored by the author and a research assistant. Inter-rater agreement for these procedures exceeded .95. Because the different cases and case variants did not contain equal numbers of propositions, analyses were performed on the percentage of propositions recalled. ANOVA was used to analyze the data in a 3 x 2 x 2 (Expertise level x Lab data format x Case) design for the manipulation of lab data, and in a 2 x 2 (Expertise level x Recall condition) design for the intentional versus incidental recall condition. The Student-Newman-Keuls test (significance level of .05) was used to make post-hoc comparisons between the subject groups.

Results

Incidental versus Intentional Recall Condition

Study time

The relationship between study time and expertise level for both intentional and incidental recall instructions is depicted in Figure 2.1. The main effect of recall condition on study time was not significant ($F(1, 42) = .003, MS_e = 1.02, p = .96$), indicating that subjects did not use more time to study the case information in the intentional recall condition than in the incidental recall condition. The main effect of level of expertise, however, was highly significant ($F(2, 42) = 106.3, MS_e = 42111, p = .0001$). The experts used about half the time to study the case than the students did.

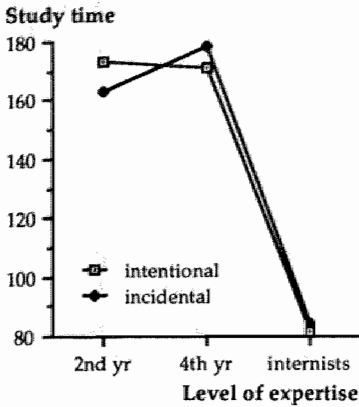


Figure 2.1. Study time as a function of expertise level and recall instructions.

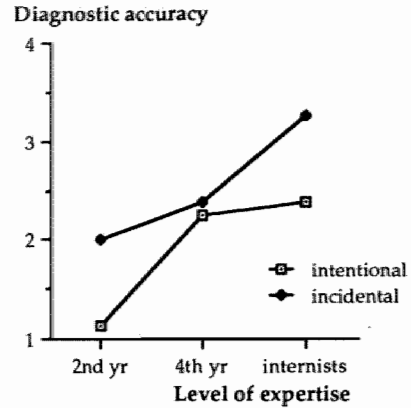


Figure 2.2. Diagnostic accuracy as a function of expertise level and recall instructions.

Diagnostic accuracy

In both the incidental and intentional recall condition subjects were asked to study a case in order to formulate a diagnosis. Therefore, no difference in diagnostic accuracy was expected between the two recall conditions. This was indeed what was found ($F(1, 42) = 1.65$, $MS_e = 4.69$, $p = .21$). However, in contrast to what might be expected, level of expertise did not show an effect on diagnostic accuracy ($F(2, 42) = 2.22$, $MS_e = 6.33$, $p = .12$). The relationship between diagnostic accuracy, level of expertise and recall condition is depicted in Figure 2.2. This figure shows a linear trend in the data, suggesting that the more experienced subjects provided better diagnoses. Polynomial contrast analysis indeed revealed a significant linear component ($F(1, 42) = 4.39$, $MS_e = 12.50$, $p = .042$), without a significant quadratic component ($F(1, 42) = .059$, $MS_e = .17$, $p = .81$). Overall the diagnostic accuracy was low, suggesting that this case of stomach carcinoma was rather difficult both for students and experts.

Recall

Whereas the subjects in the incidental recall condition were not aware that they would be asked to recall the case, the subjects in the intentional recall condition were told that they would be asked to write down whatever they remembered from the case after they formulated a diagnosis. Figure 2.3 depicts the relationship between the percentage of propositions recalled, level of expertise and recall condition. Although the picture suggests that subjects in the intentional recall condition provide more elaborate recall protocols than subjects in the incidental recall condition, the main effect of recall condition on the percentage of propositions recalled was not significant ($F(1, 42) = 2.66$, $MS_e = 54.2$, $p = .11$). The main effect of expertise level on percentage of propositions recalled was significant ($F(2, 42) = 16.3$, $MS_e = 332.0$, $p =$

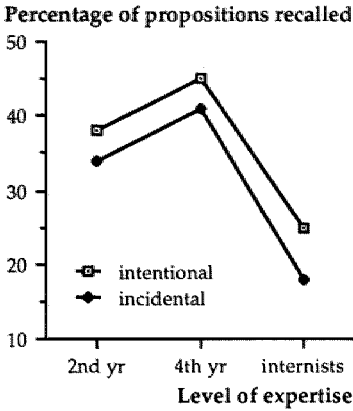


Figure 2.3. Percentage of propositions recalled as a function of expertise level and recall instructions.

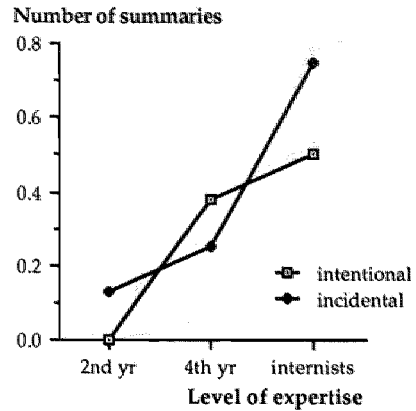


Figure 2.4. Number of summaries produced as a function of expertise level and recall instructions.

.0001). Both student groups provided more elaborate recall protocols than the internists. These results are consistent with the findings of Schmidt and Boshuizen (1993b) and suggest the existence of an intermediate effect in the recall of this case.

Summaries

Summaries were defined as inferences referring to two or more propositions in a case description. Analysis of the number of summaries produced in the recall protocols revealed no significant main effect for recall condition ($F(1, 42) = .30$, $MS_e = .083$, $p = .59$) and a significant main effect for level of expertise ($F(2, 42) = 4.54$, $MS_e = 1.27$, $p = .016$). Figure 2.4 shows that in both the incidental and intentional recall condition subjects of higher levels of expertise produced more summaries in their recall protocols. Polynomial contrast analysis revealed that the linear trend in the data was significant ($F(1, 42) = 9.05$, $MS_e = 2.53$, $p = .0044$), without a quadratic component ($F(1, 42) = .037$, $MS_e = .010$, $p = .85$). This finding is again in line with the Schmidt and Boshuizen (1993b) study, and supports the hypothesis that experts recall more case information in encapsulated mode than students.

Numerical versus Interpreted Lab Data Condition

Study time

Figure 2.5 shows study time as a function of lab data format and expertise level. The main effect of lab data format was not significant ($F(1, 84) = .041$, $MS_e = 35.0$, $p = .84$), indicating that the subjects did not use more time to study cases with numerical lab data than cases with interpreted lab data. The main effect of expertise, however, was highly significant ($F(2, 84) = 71.42$,

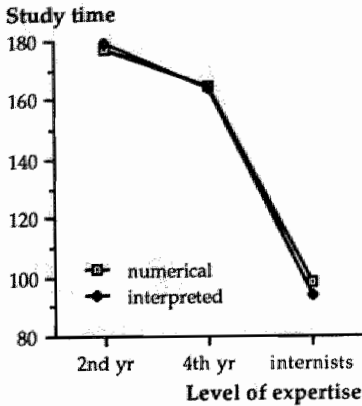


Figure 2.5. Study time as a function of expertise level and lab data format.

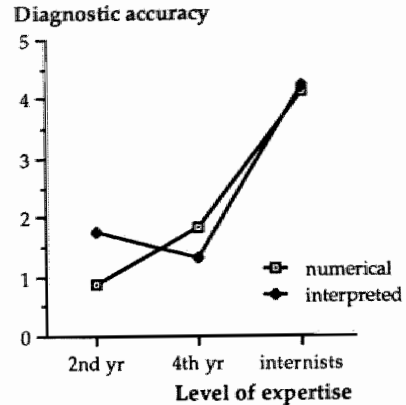


Figure 2.6. Diagnostic accuracy as a function of expertise level and lab data format.

$MS_e = 61690$, $p = .0001$): The experts used considerably less time than the students to study the cases. The main effect of cases, finally, was not significant ($F(1, 84) = .33$, $MS_e = 287.0$, $p = .57$).

Diagnostic accuracy

Lab data format had no effect on diagnostic accuracy ($F(1, 84) = .32$, $MS_e = .67$, $p = .57$). The subjects provided the same quality of diagnoses, no matter if cases were presented with numerical or interpreted lab data. Expertise level had a significant effect on diagnostic accuracy ($F(2, 84) = 81.2$, $MS_e = 39.4$, $p = .0001$). As might be expected, the experts provided much better diagnoses than both student groups (Figure 2.6). Furthermore, there was a significant case effect ($F(1, 84) = 15.8$, $MS_e = 32.7$, $p = .0001$) and a significant interaction between cases and expertise level ($F(2, 84) = 3.17$, $MS_e = 6.54$, $p = .047$). Whereas the experts diagnosed both cases equally well, the students provided better diagnoses for the heart failure case than for the pheochromocytoma case. In fact, none of the students came up with a remotely accurate diagnosis in the latter case.

Recall

The relationship between percentage of propositions recalled, level of expertise and lab data format is depicted in Figure 2.7. The main effect of lab data format on recall performance was not significant ($F(1, 84) = 1.36$, $MS_e = 0.026$, $p = .25$). This suggests that the subjects represented the clinical cases as elaborately in the numerical and interpreted lab data conditions. The main effect of level of expertise was significant ($F(2, 84) = 4.45$, $MS_e = 0.086$, $p = .015$). Pairwise comparisons between the three expertise groups revealed that the internists provided less elaborate recall protocols than both student groups. This finding is again in line with the study of Schmidt and Bos-

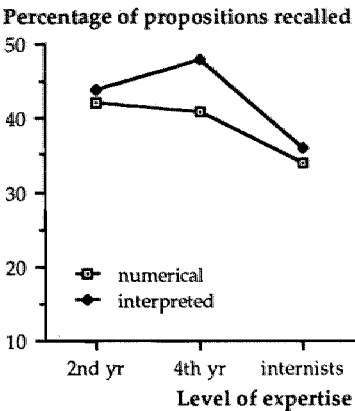


Figure 2.7. Percentage of propositions recalled as a function of expertise level and lab data format.

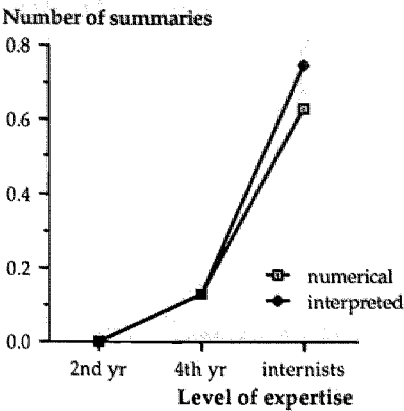


Figure 2.8. Number of summaries produced as a function of expertise level and lab data format.

huizen (1993b), suggesting an intermediate effect in clinical case recall. In addition, a significant case effect was found ($F(1, 84) = 5.98, MS_e = 0.12, p = .017$): Subjects remembered relatively more from the pheochromocytoma case than from the heart failure case.

Summaries

Analysis of the number of summaries produced in the recall protocols revealed that the main effect of lab data format was not significant ($F(1, 84) = .11, MS_e = .042, p = .74$). Thus, subjects produced similar numbers of summaries in both lab data conditions. The main effect of expertise level, on the other hand, was significant ($F(2, 84) = 11.35, MS_e = 4.29, p = .0001$). Pairwise comparisons between the expertise groups revealed that the experts produced more summaries in their recall protocols than the students (Figure 2.8). These results are again consistent with the findings of Schmidt and Boshuizen (1993b), suggesting that experts interpret clinical cases at a more encapsulated level than students. Furthermore, a significant case effect was found ($F(1, 84) = 3.97, MS_e = 1.50, p = .050$), as well as a significant interaction between cases and expertise level ($F(1, 84) = 1.50, MS_e = 3.97, p = .023$). Whereas the second- and fourth-year students hardly summarized any propositions in their recall of both clinical cases, the experts produced more summaries in the case of heart failure than in the case of pheochromocytoma.

Discussion

Since the seminal studies by de Groot (1946) on the nature of expertise in chess, free recall of the elements of a problem situation has been frequently

used as a measure of the elaborateness of the problem representation. The rationale behind the use of this measure is quite straightforward: When a person engages in the solution of a problem, he will activate relevant prior knowledge that interacts with information provided by the problem itself. This interaction produces a mental representation of the problem. The more prior knowledge the person has, that is the more expert he is with regard to the problem at hand, the more elaborate the problem representation will tend to be. If the person is subsequently asked to recall the problem information, he "reads out" the representation of it, stored in long-term memory. Therefore, the more elaborate one's prior knowledge, the more extensive the free recall. Research on expertise in a number of domains, including chess (de Groot, 1965; Chase & Simon, 1973), baseball (Spilich, Vesonder, Chiesi & Voss, 1979), bridge (Charness, 1979), electronic engineering (Egan & Schwartz, 1979), and computer-programming (McKeithen, Reitman, Rueter & Hirtle, 1981), supports this thesis. Quite different, however, seems to be the situation in the domain of medicine. In a meta-analysis of eleven clinical case recall studies, conducted between 1979 and 1989, Boshuizen has demonstrated that the relationship between level of expertise and number of idea units or propositions recalled can best be described by an inversely U-shaped curve. In other words, after an initial increase, the amount of recall actually decreases, beyond the level of intermediate subjects. Experts' recall of a clinical case is much more condensed than the recall of intermediates of the same case. Schmidt et al. (1988) have termed this phenomenon the intermediate effect in clinical case recall.⁵

Schmidt and Boshuizen (1993b) suggest that this phenomenon can be explained by assuming that the knowledge upon which experts operate becomes encapsulated through practice. Encapsulation is the subsumption or "packaging" of detailed explanatory concepts in an associative net in a smaller number of higher-level concepts (descriptions of syndromes, pathophysiological states, diagnostic labels) of greater generality.⁶ These higher-level concepts, once activated, cause experts to perceive the elements of the problem in greater "chunks" for which they have summarizing descriptions that comprise the encapsulated concepts with which they interpreted the problem in the first place. The process is considered to be largely automatic, at least in routine cases, and has a pattern-recognition-like quality. The resulting recall produced by experts is, therefore, more condensed than recall

⁵ Intermediate effects are almost exclusively found in studies in which time available to process a case was free or sufficiently long to study the material thoroughly. Schmidt and Boshuizen (1993b) have shown that, if processing time is short—in the order of 30 sec for a 250 words case description—experts show better recall than intermediates. In fact, experts' recall was demonstrated to be unaffected by manipulation of study time, whereas the performance of intermediates was greatly affected.

⁶ Knowledge encapsulation is similar to the process of compilation described by Anderson (1987). Compilation, however, operates upon procedural knowledge, whereas encapsulation is a cognitive process operating upon declarative knowledge (Schmidt & Boshuizen, 1993b).

by intermediates, who have to interpret the elements of a problem in a more effortful fashion, simply because their knowledge is of a more detailed nature, not yet tailored to the task at hand, and, hence less easily activated.

Remarkably, however, Norman and his associates (1989) have published findings that seem to be at variance with this account of expertise and its development. They presented their subjects with protocols containing numerical values of 20 laboratory data. Subjects were instructed to study a protocol and to recall the lab data as accurately as possible. The results indicated that the number of lab data recalled increased with level of expertise. In a second study, subjects either were instructed to diagnose the protocols, after which they were required to recall them as well--the authors call this the incidental recall condition--or they were instructed to memorize the material and recall it--the authors call this the intentional recall condition. In the intentional recall condition no differences emerged between students and physicians; in the incidental recall condition, however, recall by experts was better. Norman et al. (1989) explain their findings by assuming that the processing of lab data is too difficult a task to proceed in an automatic, pattern matching fashion. To make sense out of the data, even experts have to make clear efforts to analyze them before arriving at a diagnosis. As a "by-product" of this analysis, their recall of the case tends to be more elaborate.

The purpose of the present study was to test an alternative explanation for the Norman et al. findings. It was suggested that their experimental task was atypical for experts, in the sense that, almost always, lab data are evaluated in the light of pre-existing diagnostic hypotheses that are based on the patient's history and physical examination. Experts do not have to construct a representation of the patient's problem based on lab data: an initial representation usually already exists and lab data are used to confirm or rule out alternative diagnostic possibilities. It was, therefore, predicted, that, if lab data were embedded in the natural context of the patient's history and physical examination data, processing would be automatic. In addition, it was predicted that under these conditions, variations in the instruction would not matter; since the construction of a problem representation in the experts' minds would be automatic, the instruction to diagnose (with incidental recall) or the instruction to memorize (with intentional recall) would not make any difference. The findings generally confirmed these predictions. When required to process cases consisting of a patient's history, a physical examination, relevant lab data and some additional findings, experts produced more accurate diagnoses than students in both intentional and incidental recall conditions. Furthermore, experts needed less time to study the case than students (suggesting that processing was fast), provided less elaborate recall protocols (suggesting that information was processed in encapsulated mode), and produced more summaries in their recall protocols (again suggesting that information was processed in encapsulated mode).

The present study provided further evidence for the automatic nature of the processing of lab data under natural conditions. Performance of subjects was compared presenting them with two versions of the same cases: cases in which lab data were presented in numerical format, to be interpreted by the subjects, and parallel cases containing lab data in interpreted format. If interpretation of numerical data required additional processing time, as suggested by Norman and colleagues, then differences in processing time and subsequent recall were to be expected. If, however, numerical data in the context of a patient's history and physical examination data are processed automatically, as Schmidt and Boshuizen would contend, then differences were not to be expected. The findings support the automatic processing point of view. Experts studying the interpreted cases did not need less time than experts studying the material in numerical format. In addition, experts used only half of the time to study the cases than students, and produced much better diagnoses for both versions of the cases. Furthermore, experts provided less elaborate recall protocols than students irrespective of lab data condition. Finally, the higher number of summarizing propositions in the recall protocols of experts in both lab data conditions supports the hypothesis that experts process cases automatically using knowledge in encapsulated form.

The present study, thus, consistently reproduced the intermediate effect in clinical case recall for three different cases and under four different experimental conditions. The stability of this effect provides support for the theory of knowledge encapsulation proposed by Schmidt and Boshuizen (1992). If, through practice, experienced physicians have encapsulated or subsumed networks of detailed biomedical knowledge into higher-level clinical knowledge, they will automatically activate the higher-level encapsulating concepts in comprehending clinical cases. In this way, medical experts perceive sets of patient data as integrated wholes representing clinical cases in a highly condensed and meaningful form. Students, by contrast, have not yet developed knowledge in encapsulated form, and in interpreting clinical case information they will elaborate on their detailed pathophysiological knowledge resulting in more elaborate case representations.

Further study should focus on the nature of the knowledge base of subjects of different levels of expertise and the changes that occur in the course of the learning process. Research on knowledge encapsulation, furthermore, need not be confined to the domain of medicine, because it may well be possible that encapsulation of declarative knowledge is involved in other domains as well.

References

- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, 94, 192-210.

- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.
- Boshuizen, H. P. A. (1989). *De ontwikkeling van medische expertise; een cognitief-psychologische benadering*. [The development of medical expertise; a cognitive psychological approach]. Doctoral dissertation. Haarlem: Thesis Publishers.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, 33 (1), 1-16.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education*, 19, 61-67.
- Coughlin, L. D., & Patel, V. J. (1987). Processing of critical information by doctors and medical students. *Journal of Medical Education*, 62, 818-828.
- De Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). The Hague: Mouton Publishers.
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, 7(2), 149-158.
- Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.
- Hassebroek, F., Johnson, P. E., Bullemer, P., Fox, P. W., & Moller, J. H. (1993). When less is more: Representation and selective memory in expert problem solving. *American Journal of Psychology*, 106 (2), 155-189.
- Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). Contextual factors in the activation of first hypotheses: Expert-novice differences. *Medical Education*, 21, 471-476.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.
- Muzzin, L. J., Norman, G. R., Feightner, J. W., & Tugwell, P. (1983). Expertise in recall of clinical protocols in two specialty areas. *Proceedings of the 22nd Conference on Research in Medical Education*, 122-128.
- Muzzin, L. J., Norman, G. R., Jacoby, L. L., Feightner, J. W., Tugwell, P., & Guyatt, G. H. (1982). Manifestations of expertise in recall of clinical protocols. *Proceedings of the 21st Conference on Research in Medical Education*, 163-168.
- Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 1166-1174.
- Norman, G. R., Jacoby, L. L., Feightner, J. W., & Campbell, E. J. M. (1979). Clinical experience and the structure of memory. *Proceedings of the 18th Conference on Research in Medical Education*, 214-219.
- Patel, V. L., & Frederiksen, C. H. (1984). *Propositional representations of medical texts by medical students and physicians*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Patel, V. L., Groen, G. J., & Frederiksen, C. H. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education*, 20, 3-9.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93-125). New York, NY: Cambridge University press.
- Patel, V. L., & Medley-Mark, V. (1986). Relationship between representation of textual information and underlying problem representation in medicine. *Cognitive research series*. Montreal: McGill University, Centre for Medical Education. (CME86-CS1)
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-282). New York: Springer, Verlag.

- Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychological Review*, 5 (3), 205-221.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, 21 (3), 338-351.
- Schmidt, H. G., Boshuizen, H. P. A., & Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 139-145.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 275-290.
- Voss, J. F., & Bisanz, G. L. (1985). Knowledge and the processing of narrative and expository texts. In B. K. Britton & J. B. Black (Eds.), *Understanding of expository text* (pp. 173-198). Hillsdale, NJ: Erlbaum.

Appendix 2.1

Case of heart failure

Numerical lab data version

A 70-year-old female is admitted into hospital because of increasing shortness of breath. History taking reveals that the patient has been very tired lately and tolerates her food badly. Sometimes she has chest pain, especially after dinner.

Physical examination shows a pale and tired woman. She has an irregular, unequal pulse of 100/min. The blood pressure is 110/70 mmHg and jugular venous pressure is elevated. The patient has widespread peripheral edema, and positive jugular venous pulsations. The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla. Lungs: at both sides rales at lung bases. Liver and spleen not palpable.

Laboratory results show an ESR of 2 mm/h (normal: < 12 mm/h), a Hemoglobin level of 10.8 mmol/l (normal: 7.5-10.0 mmol/l) and a PCV of 0.54 (normal: 0.36-0.47). Electrolytes normal. Creatinine 85 μ mol/l (normal: 53-97 μ mol/l), CPK 40 U/l (normal: 40-200 U/l). pH is 7.50 (normal: 7.35-7.45), PO₂ 11.6 kPa (normal: 8.7-13.1 kPa), pCO₂ 3.6 kPa (normal: 4.5-5.9 kPa), HCO₃⁻ concentration 21 mmol/l (normal: 22-28 mmol/l) and O₂-saturation 97% (normal: 93-98%).

The thoracic X-ray shows congestion of the lungs and an enlarged heart. Echocardiography shows an enlarged left atrium and ventricle. And ECG reveals atrial fibrillation.

Interpreted lab data version

A 70-year-old female is admitted into hospital because of increasing shortness of breath. History taking reveals that the patient has been very tired lately and tolerates her food badly. Sometimes she has chest pain, especially after dinner.

Physical examination shows a pale and tired woman. She has an irregular, unequal pulse of 100/min. The blood pressure is 110/70 mmHg and jugular venous pressure is elevated. The patient has widespread peripheral edema, and positive jugular venous pulsations. The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla. Lungs: at both sides rales at lung bases. Liver and spleen not palpable.

Laboratory results show a low normal erythrocyte sinking rate (ESR), a slightly increased Hemoglobin level and a slightly increased packet cell volume (PCV). Electrolytes normal. Kidney function normal. CPK low normal. Analysis of blood gasses shows metabolic compensation for respiratory alkalosis.

The thoracic X-ray shows congestion of the lungs and an enlarged heart. Echocardiography shows an enlarged left atrium and ventricle. And ECG reveals atrial fibrillation.

Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians¹

The present study was performed to investigate the qualitative changes that occur in the structure of knowledge in acquiring medical expertise. Therefore, the representation of pathophysiological knowledge was compared in subjects at four different levels of expertise. Subjects studied four clinical cases under three different time constraints, and provided a diagnosis and a pathophysiological explanation for the signs and symptoms in each case. Both diagnostic accuracy and quality of explanations increased with level of expertise. The explanations of experts, however, were less elaborate and less detailed than those of students. Constraining processing time affected the quality of explanations of advanced students, but not that of experts; conversely, the elaborateness and level of detail of explanations was affected in experts but not in students. The results are interpreted as support for the hypothesis that experts apply knowledge in encapsulated form while comprehending clinical cases, whereas students apply elaborate pathophysiological knowledge.

An old lady suffering from a heart problem has just visited her cardiologist for a check up. Her heart problem is adequately controlled by the medicines she takes and her condition is stable. For several years now she gets easily tired and has been advised to rest frequently and not to exert herself too much. Sometimes she is short of breath. In the evening her ankles are swollen. She knows perfectly how to manage her problems and which medicines to take, but actually wonders how her complaints relate to the functioning

¹ Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians (manuscript under review). Parts of this chapter appeared as Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (1994). *Differences in knowledge application by students and medical experts in clinical reasoning*. Paper presented at the Annual meeting of the American Educational Research Association. New Orleans, LA, April 4-8. (ERIC Document Reproduction Service No. ED367709)

of her heart, and decides to ask her grandchildren who attend medical school. Her grandson who is in his first year was surprised by her question. He took some time to think and said: "If the heart does not work well, it cannot pump the blood adequately into the blood vessels. I think that some blood gets stuck in the legs giving you swollen ankles. But I don't know why you become easily tired and short of breath, I have not studied the subject yet." The next week the old lady asks her granddaughter who is a clerk in the hospital to explain her complaints: "The heart pumps blood throughout the circulatory system. Blood transports oxygen and nutrients to the organs and tissues. If the heart is not functioning well, some tissues and organs will receive less oxygenated blood. The extremities then become cold and pale, and you feel tired because the perfusion of your skeletal muscles is reduced. The blood flow will also be reduced in your kidneys, but not in your brain and the muscles of your heart. This is automatically regulated through vasoconstriction of your blood vessels, one of the compensatory mechanisms in heart failure. There is also a hormonal mechanism, the angiotensin-aldosterone system, that stimulates water and salt retention by the kidneys. Therefore the blood volume will increase and, therefore, I think the blood pressure. Fluid extravasates from the blood vessels in the ankles to the intracellular space because the hydrostatic pressure is greater than the osmotic pressure. So you get swollen ankles which dimple when pressure is applied and it is, therefore, called pitting edema. In the lungs you have also edema, resulting from ...". Then her grandma interrupts saying that she understood some, but now lost the thread because it was really too difficult for her. The next time when she sees her physician she asks him to explain her complaints: "It is a common problem in the elderly that the heart functions less well. The heart pumps the blood less adequately, and you notice this if you have cold and pale fingertips. Your muscles do not receive so much blood either and, therefore, you are feeling easily tired. Some blood remains in the heart causing an accumulation of fluid in the lungs. This results in less oxygen being absorbed in the blood, and you notice this on exertion, for example when you walk up the stairs and get short of breath. Furthermore, the return of blood to the heart deteriorates causing swollen ankles. This is worse in the evening because of the upright position during the day, and you know that when you put your feet on a footstool the blood will flow back from your ankles to the heart more easily." The old lady nods and says she understands her problems better now.

This example illustrates a well-known phenomenon in research on the development of medical expertise. The beginning student gave just a very short explanation of the complaints of his grandma, because he did not know more about her problem. The advanced student has learned much more in medical school, and she used this knowledge to provide a detailed explanation of the mechanisms underlying the complaints of her grandma.

The explanation was so detailed, however, that her grandma had difficulty understanding it. The physician, finally, gave a fairly short but complete explanation which satisfied his patient. Although the explanation of the physician was shorter than that of the advanced student, it was more comprehensible. The physician focused more on the complaints of the old lady and the biomedical concepts he used were at a different level of detail. Biomedical knowledge is a general term for all knowledge about the functioning and malfunctioning of the human body. Knowledge that is concerned with the classification and treatment of disease, in contrast, is termed clinical knowledge.² The advanced student focused much more on biomedical than on clinical knowledge. She lost herself, and her grandma as well, in a detailed description of the pathophysiological mechanisms underlying the manifestations of the disease. The physician, in contrast, used as much biomedical as clinical knowledge.

This phenomenon that the use of biomedical concepts in reasoning about medical problems first increases when students pass through medical school, and subsequently decreases when they gain clinical experience as clerks and physicians was demonstrated in various cognitive psychological studies (Bordage & Lemieux, 1991; Boshuizen, 1994; Boshuizen & Schmidt, 1992; Patel, Evans & Kaufman, 1990; Patel & Groen, 1986; Schmidt & Boshuizen, 1993b). It was generally assumed in these studies that the knowledge applied in diagnosing or explaining a medical problem reflected the underlying structure of a subject's knowledge base. The differences in the use of knowledge between students and physicians, therefore, were supposed to be due to structural changes in the knowledge base in the development from novice to expert (Bordage, 1994; Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992).

The first research that clearly demonstrated that experts predominantly used clinical and to a lesser extent biomedical knowledge in diagnosing a medical problem was a study by Patel and Groen (1986) on the reasoning strategies of experts in medical problem solving. The investigators asked cardiologists to read a description of a case of bacterial endocarditis, to recall the information in the case, to explain the underlying pathophysiology of the case³, and to formulate a diagnosis. Most of the causal rules applied in the explanations directly linked clinical data to diagnostic hypotheses without reference to biomedical knowledge. This suggests that in the knowledge

² Biomedical knowledge incorporates basic sciences, such as anatomy, physiology, biochemistry, pathology, and pharmacology. Clinical knowledge includes knowledge about factors predisposing disease, the variability of signs and symptoms associated with disease in patients, the course of disease, and methods for examination and management of disease.

³ The pathophysiological explanation methodology was introduced by Feltovich and Barrows (1984) to investigate how clinicians represent the relation between causes and consequences of disease in a patient. Pathophysiology is the medical discipline concerned with the changes that occur in the functioning of the human body as a consequence of disease.

base of experts, clinical data and diagnostic hypotheses are strongly associated.

Patel, Evans and Kaufman (1990) investigated the role of biomedical knowledge in clinical reasoning by medical students. First-, second-, and fourth-year medical students⁴ were also requested to study, recall, explain and diagnose the case of bacterial endocarditis. After they had diagnosed the case, they were instructed to study three biomedical texts relevant to this case and to provide a new pathophysiological explanation. The explanations generated in diagnosing the case showed that the first-year students mostly used experiential knowledge and frequently made incorrect inferences. The second-year students provided better explanations using as much biomedical as clinical knowledge. The fourth-year students, who already had two years of clinical experience as clerks, provided the most accurate explanations and diagnostic conclusions using more clinical than biomedical knowledge. The explanations generated after having studied the biomedical texts showed that all subjects used the information in the texts to update their explanations and applied more biomedical than clinical knowledge. The explanations of the fourth-year students, however, were more coherent than those of the first- and second-year students. Whereas the fourth-year students provided detailed explanations for the clinical data in relation to the diagnostic hypotheses, the first- and second-year students focused in depth on some of the signs and symptoms in the case and made more inaccurate inferences. These results indicate that the fourth-year students were best able to integrate the information in the biomedical texts with the biomedical and clinical knowledge they had acquired in the course of their medical training.

To further explore the role of biomedical knowledge in the diagnosis of clinical cases, Boshuizen and Schmidt (1992) performed an on-line study of diagnostic reasoning in which subjects of four levels of expertise participated. Second-, fourth-, and fifth-year medical students⁵ and family physicians were presented a case of pancreatitis on a sequence of cards. Subjects were instructed to think aloud in proceeding through the case and to provide a differential diagnosis at the end. The fourth-year students applied the largest number of biomedical concepts in their think-aloud protocols, whereas the fifth-year students and family physicians applied least. The same pattern was found when the use of biomedical knowledge was expressed as the proportion of biomedical concepts in the think-aloud protocols. Boshuizen and Schmidt hypothesized that the relative absence of biomedical knowledge application in the protocols of experienced subjects was the result of struc-

⁴ The undergraduate curriculum in most North American medical schools consists of a four-year program.

⁵ In contrast to the situation in North America the undergraduate medical curriculum in the Netherlands consists of a six-year program with clinical clerkships in the fifth and sixth year.

tural changes in the knowledge base with clinical experience. They supposed that students in the course of their medical training build up large networks of biomedical knowledge. Under the influence of practical experience in diagnosing and treating patients as clerks and physicians, the knowledge base is restructured in such a way that biomedical knowledge becomes integrated in clinical knowledge. Networks of detailed biomedical concepts become subsumed under a smaller number of more comprehensive clinical concepts or diagnostic labels. Schmidt and Boshuizen (1992) termed this process of structural change *knowledge encapsulation*.

In another study Schmidt and Boshuizen (1993b) used the same experimental tasks as Patel and Groen (1986), but now with a full range of expertise levels. Health sciences students, second-, fourth-, and sixth-year medical students, and internists were presented with the case of bacterial endocarditis. Subjects were requested to study the case in order to formulate a diagnosis, to recall the case and to provide a pathophysiological explanation of the signs and symptoms in the case. The results showed an inverted U-shaped relationship between the elaborateness of the pathophysiological explanations and level of expertise: The fourth- and sixth-year students were more elaborate in their explanations than either the novices, beginning students and experts. The investigators also compared the explanations of the subjects with a canonical explanation of the endocarditis case. This canonical explanation contained a minimal but sufficient set of biomedical and clinical knowledge which causally explained the signs and symptoms in the case. Most of the explanatory concepts in these explanations were of a high level of encapsulation. The comparison showed that the explanations of the more experienced subjects had more concepts in common with the explanatory concepts in the canonical explanation. Thus, although the experts provided shorter explanations of the case, these were qualitatively better. Schmidt and Boshuizen interpreted these results as support for their notion of knowledge encapsulation: Because the knowledge of experts was structured in encapsulated form, they could explain the case applying the high-level explanatory concepts which were specified in the canonical explanation and which encapsulated detailed pathophysiological mechanisms. The advanced students, on the other hand, elaborated on the pathophysiological knowledge they had acquired during medical training to provide a diagnosis and coherent explanation of the case. The novices, finally, obviously had a lack of relevant biomedical and clinical knowledge. Schmidt and Boshuizen further tested the notion of knowledge encapsulation by constraining the time available to study the clinical case. They assumed that for medical students case comprehension is an effortful, conscious process, activating potentially relevant biomedical knowledge. Experts, on the other hand, were supposed to activate relevant clinical knowledge immediately at an encapsulated level in case comprehension. Obviously, the activation of knowledge takes time. Schmidt and Boshuizen, therefore, supposed that

constraining study time would impede the processing of the case by students but not by experts. In the pathophysiological explanations this would be reflected in less elaborate and less adequate explanations for students under shorter processing time conditions. The experts' explanations were expected to be as elaborate and accurate under all processing time conditions. For the novices and the beginning students it was indeed found that the explanations were less elaborate when processing time dropped from 3 min 30 sec to 1 min 15 sec, and subsequently to 30 sec. For the advanced students the explanations were less elaborate when processing time dropped from 1 min 15 sec to 30 sec, and not when processing time dropped from 3 min 30 sec to 1 min 15 sec. The experts' explanations were indeed as elaborate under all processing time conditions. The quality of the explanations, on the other hand, seemed not only to decrease with shorter processing times for the students, but also for the experts. Schmidt and Boshuizen, however, only reported a general effect of processing time on the quality of explanations and did not spell out the effects per group.

In two on-line studies of diagnostic reasoning Bordage and Lemieux (1991) analyzed the thinking processes of students and physicians in terms of elaborateness and quality of the verbal protocols. In one study, second year students and neurologists were presented three cases of neurology, and in the other study students and gastroenterologists were presented four cases of gastroenterology. Case information was presented in segments and subjects were requested to think-aloud while proceeding through the cases. The quality of the thinking process was measured in terms of diagnostic accuracy and correct interpretation of signs and symptoms in a differential diagnostic process. The results of this experiment showed that the experts who had the highest scores on diagnostic accuracy produced short but high quality think-aloud protocols. This type of short and meaningful protocols was referred to as compiled or encapsulated (Bordage, 1994). The other experts and some students also generated high quality think-aloud protocols but used much more words to formulate correct diagnoses. Bordage (1994) referred to this type of protocols as elaborated. Most other students diagnosed the cases incorrectly and they produced low quality think aloud protocols which were either brief or lengthy.

Boshuizen (1994) combined a cross-sectional and longitudinal methodology to investigate knowledge encapsulation: Advanced students participated in two consecutive years of their medical training, so that the performance of fourth-, fifth- and sixth-year students could be compared. Subjects were requested to study, diagnose and pathophysiologically explain a case of anemia and a case of menstrual disorder. It was found that the elaborateness of the explanations decreased with experience level. Like Schmidt and Boshuizen (1993b), Boshuizen matched the subjects' explanations to canonical explanations. However, in addition to the common concepts in these explanations, she also considered the concepts at a more detailed level than the

canonical explanations and the concepts which had no relation with the concepts in the canonical explanations. The analysis showed that the number of common concepts was not higher for the more experienced students, as was found in the study of Schmidt and Boshuizen (1993b); the number of detailed concepts, however, decreased with level of experience, and so did the number of unrelated concepts. Because the more experienced students provided shorter explanations, which contained less detailed biomedical concepts and which were more confined to the canonical explanations, Boshuizen concluded that encapsulation of biomedical knowledge already starts in the period of the clerkships.

The studies described above invariably showed that beginning students missed the knowledge necessary to adequately diagnose and explain clinical cases. Intermediate students, on the other hand, who in the course of their medical training have built up a large base of detailed biomedical knowledge clearly applied this knowledge in the experimental tasks. These students usually activated all they knew about a subject, even if the knowledge was irrelevant, which resulted in lengthy think-aloud protocols and pathophysiological explanations. Furthermore, they frequently made wrong inferences and mistakes in their diagnostic conclusions. Finally, experts hardly applied any biomedical knowledge in diagnostic thinking and provided rather concise and summarizing pathophysiological explanations. Nevertheless, the experts provided the most accurate diagnoses, supporting the general assumption that experts developed the most complete and adequate set of medical knowledge throughout the years of practicing clinical medicine.

An important conclusion to be drawn from these results is that the elaborate knowledge of experts, in contrast to that of students, seems to be organized in such a way that it is readily applicable in practical settings. The encapsulation of biomedical knowledge as proposed by Schmidt and Boshuizen (1992), and adopted by Bordage (1994), is a form of knowledge reorganization that enables physicians to apply their knowledge efficiently to the task at hand. In this theory, it is supposed that the large base of biomedical knowledge students acquire in the course of their medical training is organized as large semantic networks. In solving medical problems students reason through these networks, activating all the knowledge that is related to the problem at hand. A good example of this kind of reasoning is the explanation of the old lady's granddaughter in the introduction of this article. Through repeated application of knowledge in diagnosing and treating patients in medical practice, elaborate networks of biomedical knowledge become encapsulated into higher level clinical knowledge. In the knowledge base of experts networks of detailed pathophysiological mechanisms relating causes and consequences of disease are, thus, captured under relevant clinical concepts or diagnostic labels. The more experience a physician has with a

certain disease, the stronger the associations between the findings and the diagnostic hypotheses. For example if a physician knows that an old lady has a heart problem and complains about cold fingers, swollen ankles and shortness of breath on exertion, he will automatically activate the diagnostic hypotheses of left-sided heart failure and pulmonary edema. These diagnostic hypotheses are considered encapsulating concepts: The concepts refer to a limited set of anatomical and causal pathophysiological knowledge including possible causes of disease and resulting signs and symptoms that could be observed in patients. Thus, when a physician concludes that a patient suffers from pulmonary edema, or explains shortness of breath as a consequence of pulmonary edema, he indirectly refers to the whole set of knowledge underlying the concept.⁶ For a physician pulmonary edema is not just a diagnostic label, but a concept that really fully and satisfactorily explains the condition of the patient.

The encapsulation of biomedical knowledge is not a permanent condition of the knowledge base of experts: Detailed biomedical knowledge can be activated any time it is required in a particular task situation. In routine diagnostic situations encapsulating concepts are directly activated allowing the rapid processing of clinical information in daily practice. In less straightforward diagnostic situations, however, when patient data do not consistently point towards to a limited set of encapsulating concepts, physicians can fall back on their detailed biomedical knowledge to comprehend a patient's disease process (Norman, Trott, Brooks & Smith, 1994; Patel, Groen & Arocha, 1990; Schmidt & Boshuizen, 1993a).

In the study of Patel, Groen & Arocha (1990), the diagnostic reasoning of medical specialists inside and outside their domain of expertise was investigated. Cardiologists and endocrinologists were asked to study, recall, explain and diagnose a cardiological case of cardiac tamponade with pleural effusion and an endocrinological case of Hashimoto's thyroiditis. Both of the case descriptions included symptoms unrelated to the main diagnosis. The cardiological case was found to be more difficult than the endocrinological case. The results showed that the endocrinologists generated very short explanations for the case in their own domain of expertise applying equally as many clinical as biomedical concepts. The explanations for the case outside their domain of expertise, however, were much longer and contained more biomedical than clinical concepts, indicating that they indeed activated detailed biomedical knowledge in thinking about this case. In explaining the difficult cardiological case, even the expert cardiologists generated quite extensive protocols in which they applied more biomedical than clinical knowledge. These experts utilized their biomedical domain knowledge to explain the

⁶ The physician who explained the complaint of shortness of breath to his patient in the introduction of this article used the more comprehensible term "accumulation of fluid in the lungs" instead of the clinical term of pulmonary edema. These terms are interchangeably used by physicians depending on whatever is appropriate for the particular situation.

symptoms which were unaccounted for by the main diagnosis but were indicative of a secondary disease.

Norman et al. (1994) used particularly complex clinical cases to study the differences in diagnostic reasoning between subjects with different levels of postgraduate training. Subjects were first-year residents in family medicine, first- and second-year residents in internal medicine, and experienced nephrologists. They were presented with nephrology cases consisting of a brief history and a standard array of laboratory data, and were asked to diagnose these cases while thinking aloud. It was found that the more experienced subjects provided better diagnoses, provided more causal pathophysiological explanations for the data, and combined individual data more often into meaningful clusters. Norman and colleagues concluded that, at least in this specific domain, experts revert to their knowledge of basic biomedical mechanisms when confronted with difficult clinical cases. The meaningful clusters suggest that the experts used *unfolded* encapsulated knowledge.

Altogether, the experimental findings suggest that in diagnosing routine clinical cases expert physicians activate knowledge in encapsulated form, whereas advanced students activate more detailed biomedical knowledge.⁷ The case findings, their interpretation and the knowledge activated in comprehending the case form the representation of the clinical problem. Thus, whereas experts can represent clinical cases in condensed form, students represent cases in elaborated form. However, it is questionable if the methods applied in the reported studies are powerful enough to actually chart the structure of the knowledge underlying the representation of cases. A measure as the elaborateness of think-aloud protocols and explanations is rather superficial with regard to the contents, especially if expressed as the number of words in a protocol as was the case in the study of Bordage and Lemieux (1991). Further specifications of the knowledge applied concerned the use of biomedical or clinical knowledge and various measures of the quality of protocols. Schmidt and Boshuizen (1993b) compared the pathophysiological explanations of their subjects with a canonical explanation. This seems a powerful method for analyzing the differences in the structure of knowledge underlying case representations between the different subject

⁷ In domains in which experts heavily rely on imaging techniques representing pathophysiological processes or patho-anatomical structures, different outcomes have been found. Studies on diagnostic performance in the perceptual task domains of radiology and ECG interpretation have shown that just the expert physicians applied highly specific biomedical knowledge to distinguish between alternative diagnoses (Gilhooly et al., in press; Lesgold et al., 1988; Lesgold, 1984), whereas the visual data used in both of these studies were not characterized as particularly difficult for expert physicians. These findings suggest that the interpretation of perceptual patient data in X-rays and ECGs does not proceed in encapsulated mode but requires a more detailed analysis. Further research is required to investigate whether elaborate processing of visual data by experts is characteristic for the perceptual domains rather than encapsulated processing.

groups. However, they only counted the number of explanatory concepts which both explanations had in common. Boshuizen (1994) used additional measures to compare subjects' explanations to a canonical explanation: She also counted more detailed concepts in the subjects' explanations, and concepts that were not related to the canonical explanation. Bordage and Lemieux (1991) expressed the quality of protocols as the number of times a subject correctly interpreted the clinical data in relation to their differential diagnostic meaning. Patel and colleagues (Patel Evans & Kaufman, 1990; Patel & Groen, 1986; Patel, Groen & Arocha, 1990), however, were the only researchers who studied the causal structure of the knowledge applied in subjects' explanations. They were mainly interested in differences in patterns of reasoning between students and physicians and, therefore, provided detailed analyses of the causal lines of reasoning leading to a diagnosis. Unfortunately, however, these analyses were only performed on a few protocol excerpts.

The time manipulation in the study of Schmidt and Boshuizen (1993b) is an interesting extension of the case representation paradigm. An important characteristic of expertise is the high speed of processing of information (e.g., Chi, Glaser & Farr, 1988; Elstein, Shulman & Sprafka, 1978; Ericsson & Smith, 1991; Norman, Rosenthal, Brooks, Allen & Muzzin, 1989). Medical experts can quickly understand and diagnose a clinical case because they immediately activate the relevant knowledge. Constraining processing time, therefore, would not affect the representation of clinical cases by experts. However, it would affect the representation of clinical cases by students because they are restricted in the activation of their detailed pathophysiological knowledge.

A final remark concerns the subjects participating in these studies: Only two studies included the full range of expertise levels from novices to expert physicians (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1993b). However, in the study of Boshuizen and Schmidt subject groups were relatively small: 4 to 6 subjects in each group. In neither of these studies was more than one clinical case presented to the subjects.

The aim of the present study was to investigate more thoroughly the representation of pathophysiological knowledge underlying clinical cases by students and physicians. The method used, combined and extended the methods used by Schmidt and Boshuizen (1993b) and Boshuizen (1994). Subjects of four different levels of expertise were instructed to study four clinical cases taken from different specialty areas in internal medicine. Three processing time conditions allowed subjects to study the case for either 3 min, 1 min 15 sec, or 30 sec. They were required to formulate a diagnosis, and to provide a pathophysiological explanation of the signs and symptoms in a case. The analysis of the protocols considered in detail the concepts used, as well as the links between these concepts.

It was hypothesized that pathophysiological knowledge becomes encapsulated with clinical experience and that this will be reflected in differences in the pathophysiological explanation of cases between the expertise groups. Furthermore, it was expected that constraining processing time would affect the explanations of students but not of experts, because students will be restricted in the activation of detailed pathophysiological knowledge, whereas experts immediately activate the relevant knowledge in encapsulated mode.

Method

Subjects

Subjects were 96 students and physicians of Maastricht University: 24 second-year, 24 fourth-year and 24 sixth-year medical students and 24 internists. The years of experience in internal medicine varied for the internists from 4 to 23 years with an average of 11.5 years. Each group of 24 was subdivided into three groups of eight who were assigned to three time constraint conditions. Subdivision was random for the students groups and balanced for the internists based on their subspecialisms. Students received a small financial compensation for their participation.

Material

The materials consisted of four booklets, each containing a description of a clinical case and two blank response sheets. Each clinical case description reported some contextual information, the complaint, findings from history taking and physical examination, the relevant laboratory data and some additional findings. The four clinical cases presented were pheochromocytoma, stomach carcinoma, heart failure and liver cirrhosis. These were based on actual patients and were presented following a standard patient presentation format. Except for pheochromocytoma, these cases are fairly common in medical practice. The case descriptions were about half a page in length and consisted of 33, 42, 43, and 35 propositions respectively. The translated text of the case of heart failure is provided in Appendix 3.1.

Procedure

Subjects were told that four cases would be presented and were instructed to study each case in order to produce a diagnosis. They were also told that they would be asked to explain the pathophysiology underlying the signs and symptoms in each case. Thus, subjects subsequently studied, diagnosed and explained each case. An example case was presented to make them familiar with the case format in this experiment and to experience the reading time allowed. Depending on the experimental condition, subjects were given the opportunity to read each case for 3 min, 1 min and 15 sec, or 30 sec. Subjects

were free to use as much time as they needed for the assignments and used the blank response sheets to write down their answers. The order of case presentation was balanced.

Analysis

The accuracy of the diagnoses was determined by assigning points to correct diagnostic elements. These points were added up. For a completely accurate diagnosis a maximum score of 6 points could be obtained. For example, the complete diagnosis for the heart failure case was "left- and right-sided heart failure based on mitral valve insufficiency and atrial fibrillation". Two points were given when the diagnosis contained the term "heart failure". One additional point was given for each of the following elements: "left-sided", "right-sided", "mitral valve insufficiency", and "atrial fibrillation".

The pathophysiological explanation protocols were rewritten as semantic networks of meaningful concepts. The semantic network representation consists of nodes and links between nodes: The nodes represent the different concepts applied in an explanation, and the links between nodes represent the relation between concepts. The links were qualified as causal (cau), conditional (cond), temporal (temp), attributional (att), locational (loc), as a specification (spec), a negation, an identity relation (iden), or a class relation (isa). The method applied was in essence similar to that used by Patel and Groen (1986), and Schmidt and Boshuizen (1993b). Figure 3.1 gives an example of a semantic network derived from a student's explanation (ppno. 4.8.2.3).

In the present study eight parameters were developed to identify knowledge encapsulation. These parameters were based on two earlier studies applying a post-hoc explanation methodology (Boshuizen, 1994; Schmidt & Boshuizen, 1993b). The first parameter refers to the elaborateness of the explanations and was expressed as the total number of different concepts applied. In addition, the total number of links between concepts was counted. The explanation protocol of the student in Figure 3.1, for example, contains 14 different concepts and 13 links between concepts. The more knowledge subjects apply in explaining the signs and symptoms in the cases, the more elaborate their explanations will be. Subjects who process case information in encapsulated form, therefore, can be expected to generate relatively concise explanation protocols. For example, the fourth-year student in Figure 3.1 explained that a valve dysfunction between the left atrium and ventricle leads to leakage causing backing up of blood in the left atrium and consequently in the pulmonary circulation giving pulmonary congestion (5 concepts, no. 2 through 6). At a more encapsulated level of explanation a subject would have said that mitral valve insufficiency causes pulmonary congestion (2 concepts).

The other six parameters of knowledge encapsulation were derived from comparisons of the subjects' explanations with canonical explanations

Pathophysiological explanation

As a consequence of heart failure (1) a dysfunction of the valve between the left atrium and ventricle has evolved (2): leakage (3). The result is that blood is flowing back from the left ventricle to the left atrium (4) and subsequently backs up in the pulmonary circulation (5). Through this congestion in the lungs (6) fluid extravasates (7), and the patient becomes short of breath (8). This pulmonary congestion (6) causes the right ventricle to work harder to pump against this higher resistance (9). This also leads to a valve dysfunction (10). So, a high pressure in the venous circulation evolves (11): ankle edema (12) and liver dysfunction (13) resulting in release of liver enzymes (14).

Semantic network

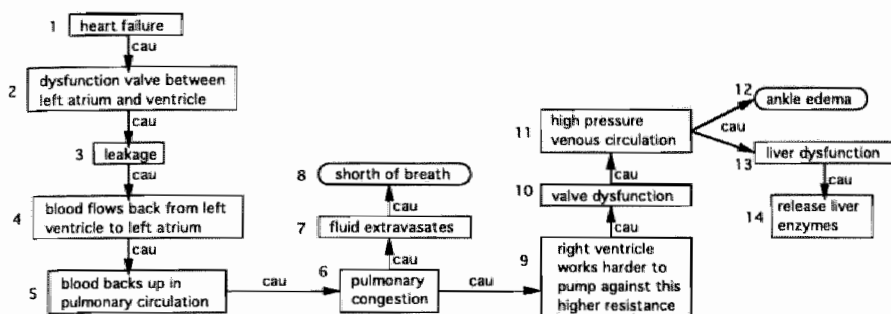


Figure 3.1. Pathophysiological explanation of the heart failure case by a fourth-year student in the 1'15" study time condition (pp 4.8.2.3), and the semantic network derived from this explanation. The different concepts used by the student are indicated by numbers. Frames with rounded corners represent signs and symptoms present in the case, and frames with rectangular corners represent explanatory concepts. The links between concepts in this explanation are qualified as causal (cau).

of the signs and symptoms in the cases. A canonical explanation of a case includes the major lines of biomedical and clinical reasoning necessary to explain all the signs and symptoms in that case. The level of detail of these explanations is just sufficient to describe the basic pathophysiological mechanisms underlying a case. The canonical models are represented as semantic networks. For each of the four clinical cases in the present experiment canonical explanations were constructed with the help of an internist and a family physician. Both experts were asked to describe the basic pathophysiological mechanisms underlying the signs and symptoms in the cases given the correct diagnoses. These descriptions were compared to each other and to the knowledge available about these diagnoses in a common textbook (Wilson, 1991). In a reiterative process, inconsistencies were solved and explanations were completed. In addition, it was attempted to establish a similar level of detail of the canonical explanations for all four cases. The number of encapsulating concepts present in the canonical explanations of these four cases considerably diverged: Whereas the canonical explanations of the cases of heart failure and liver cirrhosis contained 12

and 9 encapsulating concepts respectively, the canonical explanations of the cases of pheochromocytoma and stomach carcinoma contained only 4 encapsulating concepts each. An example of the canonical explanation of heart failure is provided in Appendix 3.2. The encapsulating concepts in this explanation are indicated by bold rectangular frames.

By matching the subjects' explanations against the canonical explanations, the quality and level of detail of subjects' explanations can be compared. The matching procedure used in the present study not only considered the concepts in an explanation, but also the links between these concepts. First, the number of concepts in a subject's explanation which were identical or equivalent to concepts in the canonical explanation was counted. These concepts were termed "model concepts". For instance, the concept of "heart failure" used by the student in our example (Figure 3.1), can be directly matched to "heart failure" in the canonical explanation, whereas the concept "dysfunction of the valve between the left atrium and ventricle" can be regarded as equivalent to "mitral valve insufficiency" in the canonical explanation, because it refers to the same pathophysiological state though in a different way (see Appendix 3.3 for a worked-out example of the matching procedure). The higher the number of model concepts applied in an explanation, the better that explanation. The quality of an explanation is expected to increase with expertise. However, if experts are able to understand a clinical case in encapsulated mode, their case explanations are not only expected to be correct, but also to be brief. A second parameter, therefore, expressed the quality of an explanation as the percentage of model concepts of the total number of concepts in that explanation. This percentage will be high if a subject's explanation is confined to the concepts used in the canonical explanation. Third, the number of concepts at a more detailed level than specified in the canonical explanation was counted (termed "detail concepts"). For example, the student in our example (Figure 3.1) used three concepts (no. 3 through 5) to explain that because of a valve dysfunction blood is flowing back from the left atrium to the pulmonary circulation causing pulmonary congestion. But in the canonical explanation a direct link from "mitral valve insufficiency" to "pulmonary congestion" is present. A high number of detail concepts is indicative for elaborate pathophysiological processing of case information, whereas a low number of detail concepts is indicative for encapsulated processing of case information.

Fourth, the number of links between concepts in the subjects' explanations which were identical to the links in the model explanation was counted (termed "model links"). An example of a model link in Appendix 3.3 is the link between "pulmonary congestion" and "edema of lungs". The higher the number of model links, the higher the concurrence with a canonical explanation, and the better a subject's explanation. A high quality explanation, however, can also be given at a more summarized level than the canonical explanation. Especially subjects who processed case informa-

tion in encapsulated form may skip some basic pathophysiological concepts and directly link clinically relevant encapsulating concepts to associated signs and symptoms in a patient. Therefore, a fifth parameter measured how often subjects abbreviated the causal line of reasoning in a canonical explanation by more directly linking model concepts by skipping intermediate nodes. These direct links were termed "short cuts". An example of a short cut in Appendix 3.3 is the direct link from "edema of lungs" to "dyspnea" which omits the decrease in oxygen level. Finally, the number of links was counted that connected model concepts in another way than was specified in the canonical explanation (termed "alternative links"). For example, the student reasoned that "heart failure" caused "mitral valve insufficiency", whereas the canonical explanation specified that "mitral valve insufficiency" caused "left-sided heart failure" (Appendix 3.3). Other examples of alternative links in Appendix 3.3 are the links in the detailed line of reasoning from "mitral valve insufficiency" to "pulmonary congestion" (concepts 2 through 6), and in the alternative line of reasoning from "pulmonary congestion" to "venous congestion" (concepts 6, 9, 10 and 11). Alternative links in an explanation, thus, may refer to incorrect lines of reasoning as well as to more detailed lines of reasoning. A high number of alternative links, therefore, can be expected in subjects who elaborately processed case information in constructing a case representation that seems plausible to them but may contain errors. A low number of alternative links, in contrast, can be expected in subjects who did not elaborate on case information, but were able to understand a clinical case immediately in encapsulated form.

Part of the data was scored by a second rater. The interrater agreement of the scoring procedures ranged from 49% for the number of detail concepts to 89% for the total number of links in the explanation protocols. The mean of the interrater agreement of all eight variables was 73%. All data were analyzed by means of repeated measures MANOVA, with the different cases as within-subjects factor, and expertise level and processing time as between-subjects factors. Polynomial contrast analysis was used to test specific hypotheses, and the Student-Newman-Keuls test (significance level of .05) was used to make post-hoc comparisons between subject groups. Pairwise comparisons between the subject groups are only reported if both a significant linear and a significant quadratic component were found in the data. Interaction effects are mostly only mentioned if they were significant.

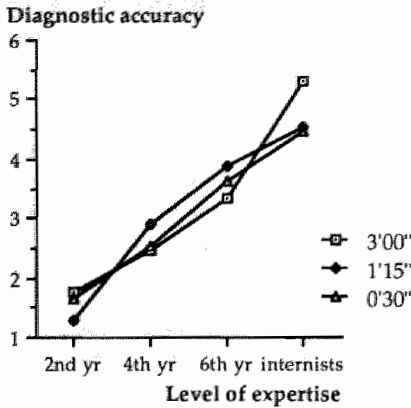


Figure 3.2. Diagnostic accuracy as a function of expertise level and processing time.

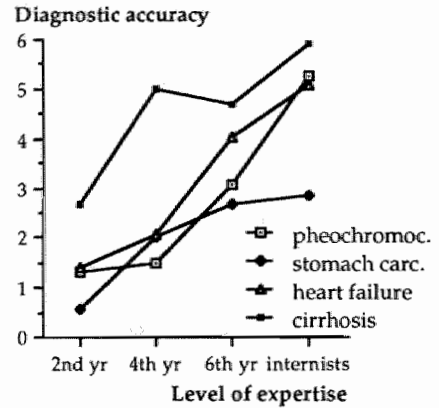


Figure 3.3. Diagnostic accuracy for each of the four cases as a function of expertise level and processing time.

Results

Diagnostic Accuracy

Figure 3.2 shows diagnostic accuracy as a function of level of expertise and processing time. A significant main effect of level of expertise on diagnostic accuracy was found ($F(3,84) = 70.59$, $MS_e = 186.5$, $p = .0001$). Polynomial contrast analysis revealed only a significant linear trend ($F(1,84) = 211.6$, $MS_e = 599.0$, $p = .0001$): Thus, the more experienced the subjects, the better diagnoses they made. The effect of processing time was not significant ($F(2,84) = .52$, $MS_e = 1.39$, $p = .59$), nor was the interaction between level of expertise and processing time. This means that for all subject groups the quality of the diagnosis is largely independent of processing time. The analysis further showed a significant case effect ($F(3,84) = 49.42$, $MS_e = 108.7$, $p = .0001$) and a significant interaction between cases and level of expertise ($F(9,84) = 5.03$, $MS_e = 11.06$, $p = .0001$). Polynomial contrast analysis for each of the four cases revealed significant linear trends of diagnostic accuracy with level of expertise for all four cases. However, some of the cases seemed to be more difficult than others (Figure 3.3). Overall the lowest scores were obtained for the case of stomach carcinoma and the highest scores were obtained for the case of liver cirrhosis, suggesting that the stomach carcinoma case was most difficult for the subjects and the liver cirrhosis case the easiest.

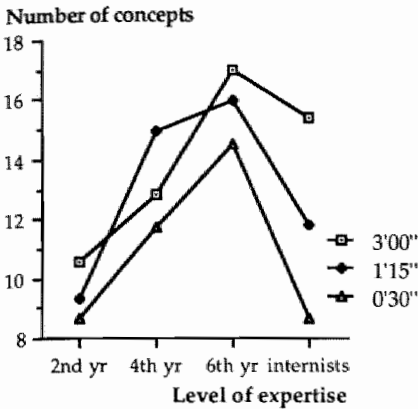


Figure 3.4. Total number of concepts in the pathophysiological explanation protocols as a function of expertise level and processing time.

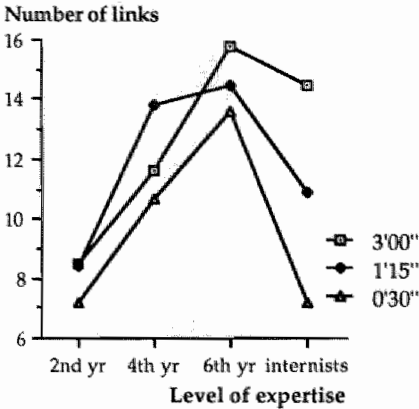


Figure 3.5 Total number of links in the pathophysiological explanation protocols as a function of expertise level and processing time.

Pathophysiological Explanations

Number of concepts

The first parameter of interest in the pathophysiological explanation protocols was the total number of concepts applied. Figure 3.4 depicts the relationship between the average number of concepts produced in the pathophysiological explanations and level of expertise. Both a significant main effect of level of expertise ($F(3,84) = 9.26, MS_e = 657.9, p = .0001$), and a significant main effect of processing time ($F(2,84) = 4.14, MS_e = 313.0, p = .015$) were found. Pairwise comparisons by the Student-Newman-Keuls test (significance level of .05) revealed that the sixth-year students used more concepts in their pathophysiological explanations than the internists, which can be interpreted as more elaborate processing by advanced students. Interestingly, constraining processing time had no significant effect on the elaborateness of the explanations of sixth-year students ($F(2,21) = .50, MS_e = 50.67, p = .61$), but had an effect on the elaborateness of the explanations of internists ($F(2,21) = 5.27, MS_e = 355.0, p = .014$). The more time the internists had to process the case information the longer their explanation protocols. Furthermore, a significant case effect was found ($F(3,84) = 22.72, MS_e = 348.0, p = .0001$).

Number of links

The average number of links between concepts in the explanations showed a similar pattern of results in relation to expertise level and processing time (Figure 3.5) as the average number of concepts. Again, significant main effects were found for level of expertise ($F(3,84) = 10.04, MS_e = 723.6, p = .0001$),

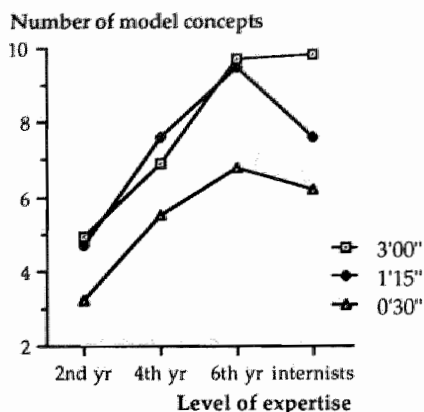


Figure 3.6. Number of model concepts in the pathophysiological explanation protocols as a function of expertise level and processing time.

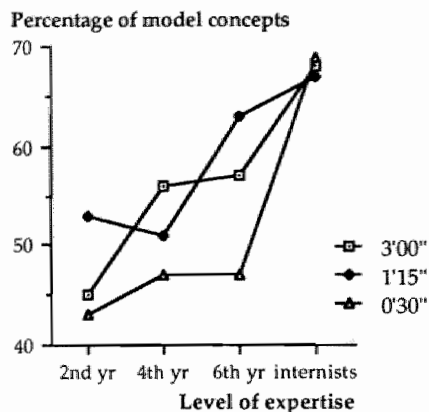


Figure 3.7. Percentage of model concepts of the total number of concepts in the pathophysiological explanation protocols as a function of expertise level and processing time.

processing time ($F(2,84) = 4.21$, $MS_e = 303.6$, $p = .018$) and cases ($F(3,84) = 19.29$, $MS_e = 310.0$, $p = .0001$). Pairwise comparisons by the Student-Newman-Keuls test (significance level of .05) also showed that the number of links increased with years of medical training for the students, and decreased for the internists. The number of links highly correlated with the number of concepts (.97). The average ratio of the number of links and the number of concepts was .87.

Number of model concepts

The comparisons of the subjects' explanations to the canonical case explanations first revealed the number of model concepts applied by the subjects. Figure 3.6 depicts the relationship between the average number of model concepts applied, level of expertise and processing time. The main effect of level of expertise was significant ($F(3,84) = 20.18$, $MS_e = 350.2$, $p = .0001$), as well as the main effect of processing time ($F(2,84) = 9.26$, $MS_e = 657.9$, $p = .0001$). Pairwise comparisons by the Student-Newman-Keuls test (significance level of .05) revealed that students with longer medical training applied more model concepts. The experts, however, did not differ from fourth- and sixth-year students concerning the number of model concepts in their explanations. All subject groups applied fewer model concepts in the shortest processing time condition. Furthermore, a significant case effect ($F(3,84) = 30.74$, $MS_e = 170.1$, $p = .0001$) and a significant interaction between cases and expertise level ($F(9,84) = 3.15$, $MS_e = 17.43$, $p = .0013$) were found.

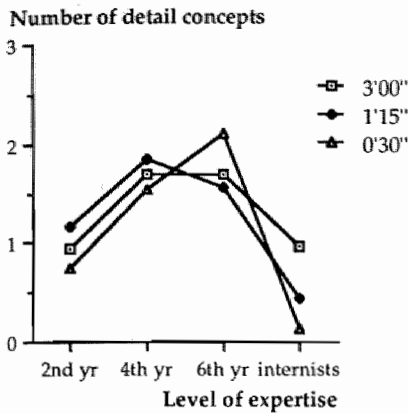


Figure 3.8. Number of detail concepts in the pathophysiological explanation protocols as a function of expertise level and processing time.

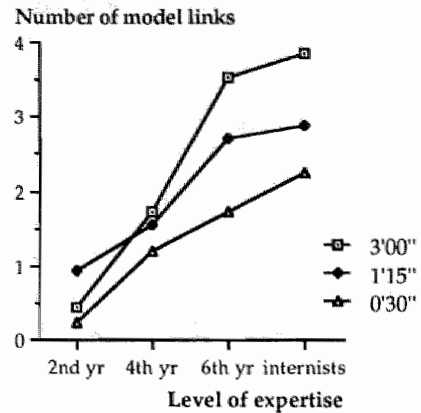


Figure 3.9. Number of model links in the pathophysiological explanation protocols as a function of expertise level and processing time.

Percentage of model concepts

The percentage of model concepts of the total number of concepts in the protocols was used as a measure of the quality of an explanation. The main effect of level of expertise on the percentage of model concepts was significant ($F(3,84) = 16.52$, $MS_e = .78$, $p = .0001$), as well as the main effect of processing time ($F(2,84) = 3.26$, $MS_e = .15$, $p = .043$). The relationship between percentage of model concepts and level of expertise (Figure 3.7) could be characterized as positively linear as indicated by a strong linear component ($F(1,84) = 45.50$, $MS_e = 2.14$, $p = .0001$) and a weak quadratic component ($F(1,84) = 3.24$, $MS_e = .15$, $p = .076$) in polynomial contrast analysis. This means that increasing levels of expertise were associated with a better match with the canonical explanations and, hence, with a higher quality of the explanations. The significant effect of processing time should be attributed to the poor quality of the students' explanations in the 30 sec condition; in this condition students had not enough time to elaborate on the case information. The percentage of model concepts in experts' explanations was *not* affected by processing time conditions ($F(2,21) = .069$, $MS_e = .004$, $p = .93$), suggesting that experts automatically activated the relevant knowledge in studying the cases under all processing time conditions. The overall case effect was significant ($F(3,84) = 6.09$, $MS_e = .16$, $p = .0001$), as was the interaction between cases and expertise level ($F(9,84) = 2.71$, $MS_e = .071$, $p = .005$).

Number of detail concepts

The number of detail concepts measured the level of detail of an explanation. Figure 3.8 shows that the relationship between number of detailed concepts and level of expertise was an inverted U-shaped one. The main ef-

fect of expertise on level of detail of explanations was significant ($F(3,84) = 6.72$, $MS_e = 35.54$, $p = .0004$), which could be specified as a significant quadratic component ($F(1,84) = 18.34$, $MS_e = 97.0$, $p = .0001$), and a non-significant linear component ($F(1,84) = 1.35$, $MS_e = 7.13$, $p = .25$). These results support the hypothesis that advanced students activate detailed biomedical knowledge in clinical reasoning, while experts do not. The main effect of processing time, however, was not significant ($F(2,84) = .26$, $MS_e = 1.25$, $p = .79$). Thus, in general, constraining processing time had no effect on the level of detail of explanations. Furthermore, a significant case effect was found ($F(3,84) = 3.92$, $MS_e = 10.70$, $p = .0092$).

Number of model links

The number of model links in the subjects' explanations and the canonical explanations is depicted as a function of level of expertise and processing time in Figure 3.9. Both a significant main effect of expertise ($F(3,84) = 21.60$, $MS_e = 122.3$, $p = .0001$), and a significant main effect of processing time ($F(2,84) = 6.38$, $MS_e = 36.11$, $p = .0026$) were found. The effect of expertise could be specified as a linear trend ($F(1,84) = 62.31$, $MS_e = 352.8$, $p = .0001$). Thus, subjects with higher levels of expertise more often connected concepts in the same manner as was specified in the canonical explanations. The number of model links also increased when subjects had more time to study the cases. In addition, a significant case effect was found ($F(3,84) = 9.52$, $MS_e = 30.27$, $p = .0001$), as was a significant interaction between cases and expertise level ($F(9,84) = 2.69$, $MS_e = 8.54$, $p = .0013$).

Number of short cuts

The number of short cuts in the reasoning of subjects, as compared to the lines of reasoning in the canonical explanations, again showed a positive linear relationship with expertise (Figure 3.10): The main effect of expertise was significant ($F(3,84) = 27.99$, $MS_e = 48.41$, $p = .0001$), with a significant linear component ($F(1,84) = 82.70$, $MS_e = 143.0$, $p = .0001$). These results indicate that more experienced subjects explained case information at a more summarized level than less experienced subjects by skipping parts of the lines of reasoning in the canonical explanation. The main effect of processing time on number of short cuts was not significant ($F(2,84) = 1.69$, $MS_e = 2.91$, $p = .19$). The case effect was significant ($F(3,84) = 14.23$, $MS_e = 20.7$, $p = .0001$), as was the interaction between cases and expertise level ($F(9,84) = 7.46$, $MS_e = 10.86$, $p = .0013$).

Number of alternative links

The number of alternative links, finally, referred to lines of reasoning not specified in the canonical explanation. The graph depicting the relationship between the number of alternative links, expertise and processing time shows an inverted U-shaped function (Figure 3.11). The main effect of ex-

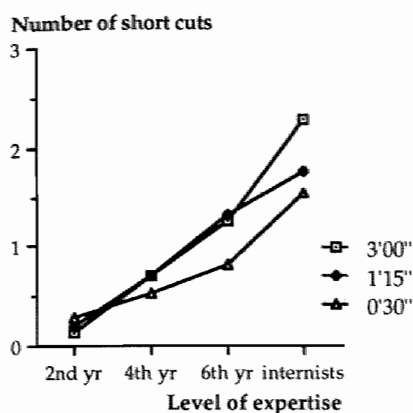


Figure 3.10. Number of short cuts in the pathophysiological explanation protocols as a function of expertise level and processing time.

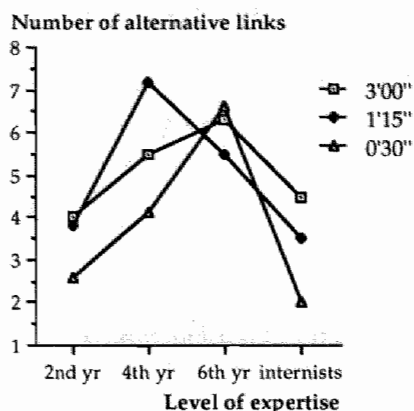


Figure 3.11. Number of alternative links in the pathophysiological explanation protocols as a function of expertise level and processing time.

pertise was significant ($F(3,84) = 7.22$, $MS_e = 198.7$, $p = .0002$), with a significant quadratic component ($F(1,84) = 21.07$, $MS_e = 580.2$, $p = .0001$) and a non-significant linear component ($F(1,84) = .006$, $MS_e = .17$, $p = .94$). The main effect of processing time on number of alternative links was not significant ($F(2,84) = 2.34$, $MS_e = 64.32$, $p = .10$), as well as the interaction effect of expertise and processing time ($F(6,84) = 1.12$, $MS_e = 30.95$, $p = .36$). These results suggest that under all processing time conditions fourth- and sixth-year students provided more alternative explanations for the case information than either beginning students or internists. The case effect was significant ($F(3,84) = 8.08$, $MS_e = 90.56$, $p = .0001$).

In combination, the data on the links between concepts in subjects' explanations clearly showed that the lines of reasoning applied by experts best corresponded to the canonical explanations. Whereas the experts tended to make short cuts in reasoning omitting concepts specified in the canonical explanations, advanced students rather set up alternative lines of reasoning in explaining the cases.

An overview of the results obtained by MANOVA is provided in Table 3.1. The effect of expertise was strong for all parameters measured, but the effect of processing time was not always present. Processing time and expertise level did not interact on any of the parameters measured, suggesting that, if an effect of processing time occurred, it was in the same direction for all subject groups. In addition, we found strong case effects for all parameters measured, and for some parameters cases and expertise level even strongly interacted. For these parameters the effect of expertise was found to be larger for the cases of pheochromocytoma and liver cirrhosis than for the cases of stomach carcinoma and heart failure.

Table 3.1. Overview of the effects of expertise, processing time, and case on diagnostic accuracy and the parameters measured in the pathophysiological explanations obtained by MANOVA (*p*-values).

	Diagnostic accuracy	Concepts	Links	Model concepts	% Model concepts	Detailed concepts	Model links	Short cuts	Alternative links
Expertise	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Processing time		< .05	< .05	< .05	< .001		< .001		
Time x Expertise									
Case	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Case x Expertise	< .001			< .001	< .001		< .001	< .001	
Case x Time									

Discussion

The basic hypothesis tested in this study was that, with experience in medical practice, the elaborate knowledge networks built up by medical students in the course of their training do not only become extended, but also become reorganized into a more readily applicable form (Boshuizen & Schmidt, 1992; Schmidt and Boshuizen, 1993b). As a result of the repeated application of knowledge in diagnosing and treating patients, the knowledge of expert physicians is supposed to be organized in such a way that clusters of detailed pathophysiological knowledge are encapsulated into more comprehensive clinical concepts or diagnostic labels. Therefore, in understanding routine clinical problems, experts will be able to immediately activate the relevant knowledge in encapsulated form, without activating the more detailed pathophysiological knowledge. Students, by contrast, who have not yet developed these encapsulated knowledge structures will activate their pathophysiological knowledge to construct a coherent representation of the problem. Basically, the encapsulation of biomedical knowledge in expert physicians thus implies that experts have developed a fairly complete and adequate set of domain knowledge, that experts can quickly apply their knowledge to the task at hand, and that they do not activate detailed biomedical knowledge if this is not necessary.

In the present study these hypotheses were tested by comparing the diagnoses and pathophysiological explanations provided for four clinical cases by students and physicians under various time constraints. The method used for analyzing the explanation protocols was more precise than in earlier studies (Boshuizen, 1994; Schmidt and Boshuizen, 1993b), and considered in detail the concepts used, as well as the links between these concepts. The results largely confirmed the hypotheses, providing additional empirical support for the theory that the knowledge of expert physicians is organized in encapsulated form whereas the knowledge of students is organized in elaborated form. On the other hand, the assumption that experts do not activate more biomedical knowledge than necessary, even if they have ample time, was not confirmed by the data. In this section the data will be reviewed in relation to the hypotheses and explanations will be provided for the discrepant findings.

If experts have developed more complete and adequate knowledge than students, it might be expected that they would provide more accurate diagnoses and qualitatively better explanations than students. This was indeed what we found. Diagnostic accuracy increased from 25% for the second-year students to 80% for the internists. The quality of the explanations was better for internists in the sense that their explanations were more confined to the canonical explanations as indicated by a higher percentage of model concepts, more model links and more short cuts in model lines of reasoning than students.

Furthermore, it was expected that the encapsulated structure of knowledge would lead experts to represent the clinical cases at a more summarized level than students. The pathophysiological explanation protocols supported this hypothesis. The explanations of experts were less extensive than those of advanced students, but at the same time they were more confined to the canonical explanations that specified the major lines of causal reasoning necessary to explain the signs and symptoms in the cases. The experts even abbreviated these lines of reasoning by skipping intermediate concepts. The fourth- and sixth-year students, on the other hand, used more detailed biomedical knowledge in their explanations than the experts, and more often applied lines of reasoning that were not specified in the canonical explanations as indicated by a higher number of alternative links. Thus, these data, which reflect the level of detail and the lines of reasoning in subjects' pathophysiological case representations more precisely than the data in earlier studies (Bordage, 1994; Bordage and Lemieux, 1991; Boshuizen, 1994; Boshuizen & Schmidt, 1992; Patel Evans & Kaufman, 1990; Patel & Groen, 1986; Schmidt and Boshuizen, 1993b), corroborate the outcomes of these studies, providing strong additional support for the hypothesis that advanced students process case information in an elaborate way by activating detailed pathophysiological knowledge, whereas expert physicians process case information at a more encapsulated level.

Moreover, the encapsulated structure of knowledge would enable experts to apply their knowledge quickly to the task at hand. Limitations of processing time, therefore, were expected not to affect the diagnoses and explanations provided by experts. Separate analyses of the expert group, however, showed that processing time had a significant effect on the task performance of experts (see Table 3.2 for a summary of the results). The two short processing time conditions resulted in identical diagnostic performance (75%), but this increased to 88% in the long processing time condition. Given longer processing time experts generated more complete diagnoses. A similar phenomenon was found regarding their pathophysiological explanations which were more elaborate when they had had more time to study the cases. These longer explanations also contained more model concepts, more detail concepts and more alternative lines of reasoning than the explanations they gave under shorter processing time conditions. However, the available processing time did *not* affect the percentage of model concepts, the number of model links, and the number of short cuts in the experts' explanations. The experts' explanations, thus, were of the same high quality in all processing time conditions, but were provided at a more summarized level of explanation when processing time was shorter. The latter findings are in line with the results on diagnostic accuracy, and clearly indicate that experts are able to activate the relevant knowledge, even when they have very little time to study the clinical cases. It seems, therefore, that experts are capable to process the case information in encap-

Table 3.2. Effects of processing time on diagnostic accuracy and pathophysiological explanations by sixth-year students and internists.

	Sixth-year students		Internists	
	Anova*	Post-hoc**	Anova	Post-hoc
Diagnostic accuracy			< .05	S vs. L M vs. L
No. of concepts			< .05	S vs. L
No. of links			< .05	S vs. L
No. of model concepts	< .10		< .05	S vs. L
% of model concepts	< .05	S vs. M		
No. of detail concepts			< .05	S vs. L
No. of model links				
No. of short cuts				
No. of alternative links			< .05	S vs. L

* *p*-values; ** sign. level $p < .05$; S: 30 sec processing time condition; M: 1 min 15 sec processing time condition; L: 3 min processing time condition.

sulated form, but elaborate on their encapsulated knowledge when they have ample time to study the cases. Therefore, the hypothesis that experts do not activate more detailed biomedical knowledge if this is not necessary to comprehend a clinical case, is not supported by the present experiment.

In contrast to experts, advanced students were supposed to apply detailed pathophysiological knowledge in comprehending clinical cases. Therefore, it was expected that they would be seriously impeded in elaborating on their knowledge when processing time was constrained, resulting in less accurate diagnoses and pathophysiological explanations. The effects of processing time on the diagnoses and the explanations of the sixth-year students are summarized in Table 3.2. Interestingly, no differences in the diagnostic accuracy of students were found in the three processing time conditions. Neither did the limitations on processing time affect the elaborateness of their explanations. These were as detailed and contained as many alternative lines of reasoning in all three processing time conditions. There seemed to be fewer model concepts in the shortest processing time condition, but this effect was not significant. The same applied to the number of model links and short cuts. However, the overall quality of the explanations, as indicated by the percentage of model concepts, appeared to decline with limitation of processing time: When students had less time to study the cases, they apparently formed qualitatively poorer case representations. In contrast to experts, 30 seconds were really too short for students to interpret and represent the relevant case information. This corroborates the hypotheses that understanding of clinical data for students is not simply a matter of automatic activation of a set of accurate clinical concepts and diagnos-

tic hypotheses, as it is for experts, but rather it is an active process of knowledge retrieval and inferencing that requires conscious effort. In explaining the pathophysiology underlying the cases, students appeared to continue this process of knowledge elaboration. However, because students' initial case representations were less adequate under shorter processing time conditions, they provided better explanations when they had had more time to study the cases. Remarkably, the better understanding of the case information under longer processing time conditions did not lead to an improvement in students' diagnostic accuracy.

The major discrepancy between the findings in the present experiment and the hypotheses formulated concerned the activation of more detailed pathophysiological knowledge by the experts in case of ample processing time. Yet, this discrepancy does not imply that the experts were not capable of processing the case information in encapsulated mode: Even when they had very little time they immediately activated the relevant knowledge. Thus, the question is why the experts activated detailed biomedical knowledge when this was not necessary to comprehend the case information.

As discussed in the introduction of this chapter, the extent to which detailed biomedical knowledge is applied is subject to variation depending on the task difficulty and the task domain. In addition, unfolding of detailed biomedical knowledge can be probed by specific task requirements (e.g., priming task in Schmidt & Boshuizen, 1993b; concept explanation in Van de Wiel, Schaper, Boshuizen & Schmidt, 1995). However, none of these task dependent explanations seemed to apply to the present study. Similar to the study of Schmidt and Boshuizen (1993b), subjects engaged in a diagnostic task and provided a pathophysiological explanation of the case information presented. Furthermore, as in all studies on clinical reasoning reviewed in this chapter, the clinical case material was presented in a standard patient presentation format consisting of the case history, the physical examination, the relevant lab data and some additional findings. Neither could the clinical cases used in the present experiment be described as particularly difficult. The most difficult case was the case of stomach carcinoma with a mean expert diagnostic accuracy of only 47%. Many experts appeared to be unable to formulate a final diagnosis on the basis of the case information presented and provided imprecise and incomplete diagnoses. The case of liver cirrhosis, on the other hand, was not at all difficult for the experts as indicated by a diagnostic accuracy score of 99%. Separate analyses for each of these cases showed in main lines the same pattern of results. The statistical effects, however, were stronger when the data of all four cases were analyzed in a repeated measures design.

Therefore, if the effect of processing time on the activation of knowledge by the expert physicians cannot be explained by the specific task conditions in the present experiment, more fundamental explanations should be explored. In order to do so, it is necessary to revisit our assumptions about

expert knowledge structure and the activation of relevant concepts. We assumed that the knowledge of experts is organized in large associative networks, and that clusters of highly connected concepts in these networks are encapsulated under more comprehensive clinical concepts. These encapsulated clusters of knowledge were supposed to consist of detailed pathophysiological mechanisms explaining the signs and symptoms of a disease process. In line with a hypothesis proposed by Schmidt and Boshuizen (1993b), we further assumed that in diagnosing routine cases experts would immediately activate the relevant encapsulating concepts, and would *not* activate the detailed pathophysiological concepts subsumed under these encapsulating concepts, even if ample time was given for case processing.

This last assumption, in fact, is neither theoretically plausible, nor is it empirically supported. Theories on associative networks (e.g., Anderson, 1983; Kintsch, 1988; Quillian, 1968) generally assume that incoming information activates corresponding concepts in the knowledge network, and that spread of activation through the network results in the activation of associated concepts. The extent of spreading activation is directly proportional to the strength of the links between these concepts and the initially activated concepts. In case of negative links, activation of the associated concepts is inhibited. Thus, the direct activation of encapsulating concepts in diagnostic reasoning by experts must reflect strong links between the activated signs and symptoms and the encapsulating concepts. In contrast, *not activating* detailed biomedical concepts, even if sufficient time is available for elaborate case processing, suggests that inhibitory links must at least exist between encapsulating concepts and the networks of detailed biomedical concepts that they subsume. This would be a rather peculiar assumption, especially, because it was also assumed that activation of detailed encapsulated knowledge is dependent on the nature of the task: Whereas experts would not activate detailed encapsulated knowledge in diagnosing routine clinical cases, they would *do* so, for example, in explaining the pathophysiological mechanisms of disease to students. Thus, if experts' knowledge can be flexibly applied to the task at hand, a mechanism would be required to set the links in an inhibitory or excitatory mode. No theory of spreading activation has this feature.

In addition, the empirical data Schmidt and Boshuizen (1993b) provided for their assumption that experts do not activate detailed concepts in routine diagnosis, even if given ample processing time, is not as convincing as suggested. Schmidt and Boshuizen (1993b) did constrain maximum processing time to either 3 min 30 sec, 1 min 15 sec, or 30 sec, but *did not measure* the time experts actually used to diagnose a case. Informal observations, however, suggested that most experts used only half of the time in the long processing time condition, presumably because they did not need more time to diagnose the case. Thus, instead of measuring the effects of an ample pro-

cessing time condition on case representation, Schmidt and Boshuizen measured the effects of a sufficient processing time condition.

Moreover, the present study suggests that the activation of detailed pathophysiological knowledge indeed depends on the available processing time. In this time manipulation study we actually controlled how much time subjects used for case processing. Experts used on average 2 min 30 sec for case processing, which was less than students used, but still considerably more than experts in the Schmidt and Boshuizen (1993b) study. We found that under extremely short processing time conditions experts used high-level encapsulating concepts and virtually no detailed pathophysiological concepts in comprehending routine clinical cases. However, when given ample processing time, experts used more of the detailed pathophysiological knowledge subsumed under the encapsulating concepts. These findings provide clear evidence that no inhibitory links exist between encapsulating concepts and the encapsulated networks of pathophysiological concepts. The encapsulated network of pathophysiological knowledge, therefore, seems not such an isolated structure that is activated or not, as conceptualized by Schmidt and Boshuizen (1992; 1993b). We would rather assume that signs and symptoms, causes, and pathophysiological mechanisms of disease are represented in one coherent knowledge network together with the relevant encapsulating concepts.

Although this is the first study that reports these effects of time constraint on experts' knowledge activation and case representation, in hindsight, the findings of Schmidt and Boshuizen (1993b) also suggested that study time has an effect on the processing of case information by experts. Additional analysis on the data set of these investigators showed that diagnostic accuracy increased for the experts from 33% in the shortest processing time condition to 71% in the longest processing time condition. Schmidt and Boshuizen (1993b) reported that this difference was probably not significant because of a lack of statistical power. It should be noted that they used only one clinical case in their experiment. Although the elaborateness of the explanations of the experts did not differ in the three processing time conditions, the number of explanatory concepts that corresponded to the canonical case explanation was higher when more time was given to study the case. This effect approached significance ($F(2,21) = 3.07$, $MS_e = 9.04$, $p = .068$), suggesting that experts were better able to activate the relevant knowledge when they had more time to study the case. It seems that 30 seconds was really too short to construct a complete representation of the fairly complex case of bacterial endocarditis that was presented.

Altogether, it seems that we cannot contend that encapsulated knowledge is *not* activated in diagnosing routine clinical cases. What is more plausible is that this detailed kind of knowledge is activated, but not strongly enough to be part of the case representation (Anderson, 1983; Kintsch, 1988) when cases are quickly processed. Activation level of the en-

capsulated knowledge is low, because in experts' knowledge networks the links between signs and symptoms, encapsulating concepts, and causes of disease are much stronger than the links between these concepts and the concepts representing the pathophysiological mechanisms of disease. When experts process a case for a longer time, the concepts in the knowledge network representing the case information will be more strongly activated and, hence, the intermediate pathophysiological concepts will be more strongly activated. In clinical practice, however, physicians usually work under a high time pressure. It seems likely that under these circumstances they would not use more time than necessary to solve a medical problem applying the most direct lines of reasoning (from signs and symptoms via encapsulating concepts to possible causes of disease). This was clearly demonstrated in studies that measured the time taken to diagnose clinical cases (e.g., Elstein et al., 1978; Norman et al., 1989): Experts were able to provide accurate diagnostic hypotheses in a very short time.

References

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Bordage, G. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Academic Medicine*, 69(11), 883-885.
- Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Academic Medicine*, 66(9), S70-S72.
- Boshuizen, H. P. A. (1994). Cognitive effects of practical experience in high- and low-achieving medical students. *Learning and Instruction*, 4, 313-329.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving*. Cambridge, MA: Harvard University Press.
- Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge: Cambridge University Press.
- Feltoovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt & M. L. De Volder (Eds.), *Tutorials in problem-based learning. New directions in training for the health professions*, (pp. 128-142). Assen/Maastricht: Van Gorcum.
- Gilhooly, K. J., McGeorge, P., Hunter, J., Rawles, J. M., Kirby, I. K., Green, C., & Wynn, V. (in press). Biomedical knowledge in diagnostic thinking: the case of electrocardiogram (ECG) interpretation. *European Journal of Cognitive Psychology*.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Lesgold, A., Robinson, H., Feltoovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill; Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*, (pp. 311-342). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory 'Essays in honor of Gordon Bower'*, (pp. 31-60). San Francisco, New York: W. H. Freeman and Company.
- Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125, 1063-1068.

- Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E. K. M. (1994). Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine*, 6(2), 114-120.
- Patel, V. L., Evans, D. A., & Kaufman, D. R. (1990). Reasoning strategies and the use of biomedical knowledge by medical students. *Medical Education*, 24, 129-136.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition*, 18(4), 394-406.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Massachusetts: MIT Press.
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-282). New York, NY: Springer Verlag.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), 205-221.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, 21(3), 338-351.
- Van de Wiel, M. W. J., Schaper, N. C., Boshuizen, H. P. A., & Schmidt, H. G. (1995). *The explanation of encapsulating concepts by medical experts, clerks and advanced students*. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, April 18-22. (ERIC Document Reproduction Service No. ED385194)
- Wilson, J. D., Braunwald, E., Isselbacher, K. J., Petersdorf, R. G., Martin, J. B., Fauci, A. S., & Root, R. K. (Eds.). (1991). *Harrison's Principles of Internal Medicine* (12th ed.). New York: McGraw-Hill.

Appendix 3.1

Case of heart failure

A 70-year-old female is admitted into hospital because of increasing shortness of breath. History taking reveals that the patient has been very tired lately and tolerates her food badly. Sometimes she has chest pain, especially after dinner.

Physical examination shows a pale and tired woman. She has an irregular, unequal pulse of 100/min. The blood pressure is 110/70 mmHg and jugular venous pressure is elevated. The patient has widespread peripheral edema, and positive jugular venous pulsations. The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla. Lungs: at both sides rales at lung bases. Liver and spleen not palpable.

Laboratory results show an ESR of 2 mm/h (normal: < 12 mm/h), a Hemoglobin level of 10.8 mmol/l (normal: 7.5-10.0 mmol/l) and a PCV of 0.54 (normal: 0.36-0.47). Electrolytes normal. Creatinine 85 μ mol/l (normal: 53-97 μ mol/l), CPK 40 U/l (normal: 40-200 U/l). pH is 7.50 (normal: 7.35-7.45), PO₂ 11.6 kPa (normal: 8.7-13.1 kPa), pCO₂ 3.6 kPa (normal: 4.5-5.9 kPa), HCO₃⁻ concentration 21 mmol/l (normal: 22-28 mmol/l) and O₂-saturation 97% (normal: 93-98%).

The thoracic X-ray shows congestion of the lungs and an enlarged heart. Echocardiography shows an enlarged left atrium and ventricle. And ECG reveals atrial fibrillation.

Appendix 3.2

Canonical explanation of the case of heart failure

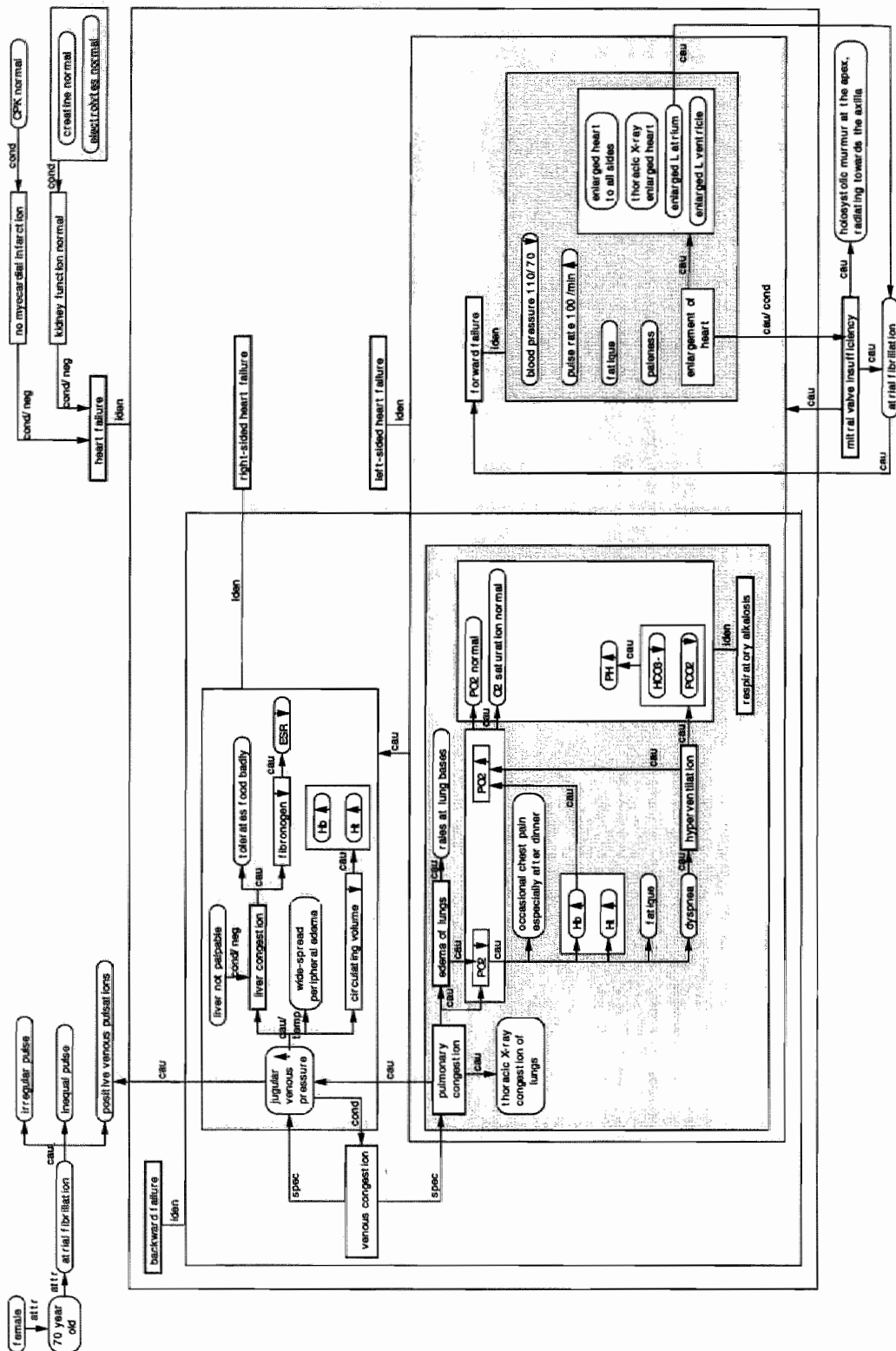
Frames with rounded corners represent signs and symptoms present in the case, and frames with rectangular corners represent explanatory concepts. Explanatory concepts that are con-

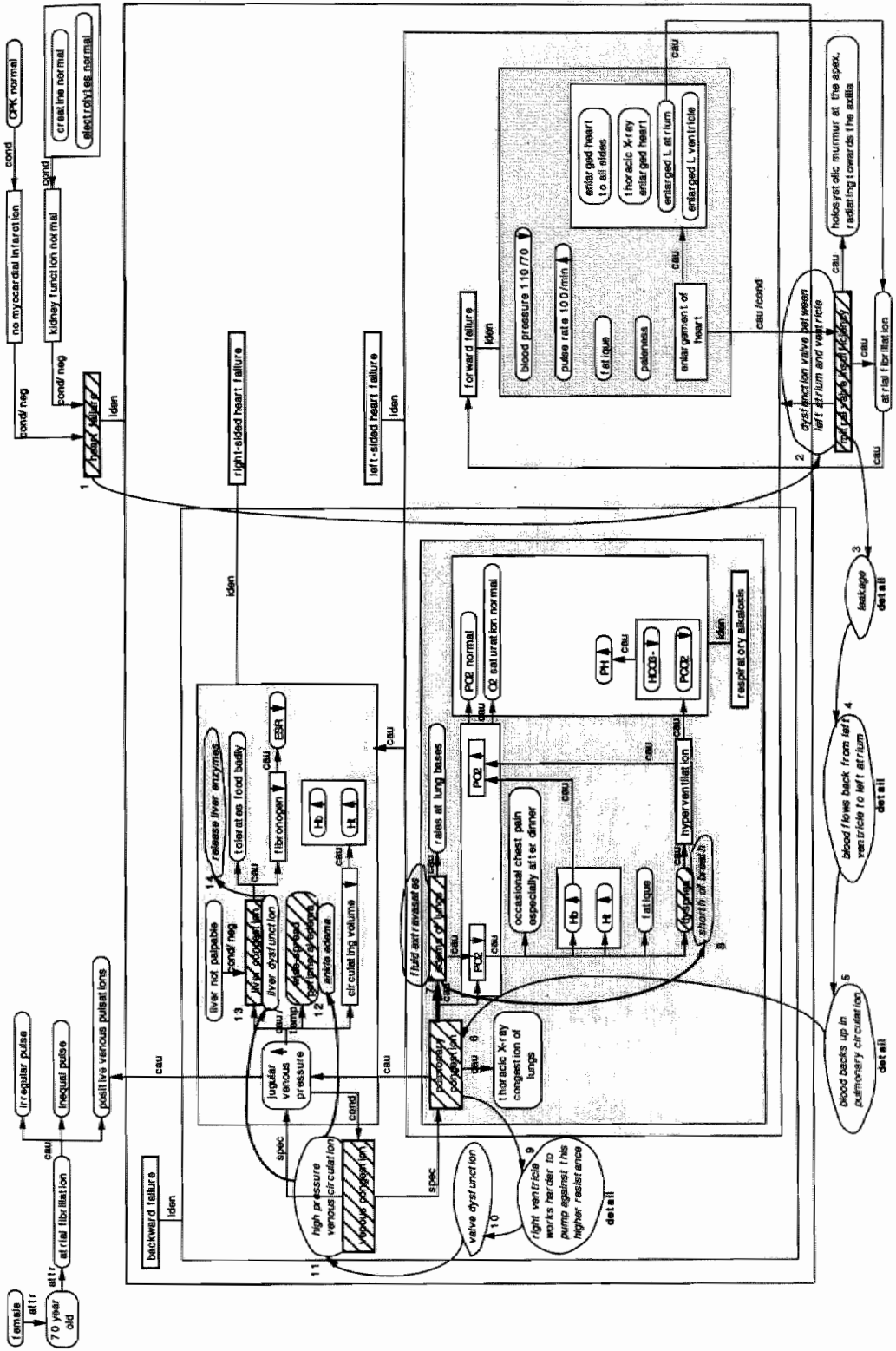
sidered encapsulating concepts have bold frames. Most links between concepts are directed, as indicated by the arrows. These links could be qualified as causal (cau), conditional (cond), attributional (attr), or temporal (temp). In addition, a concept could be a specification of another concept (spec), or a concept could contradict another concept (cond/neg). Identity relations between concepts (iden) are non-directed; these relations mostly imply that one concept can be used to refer to a complex of findings. Boxes around a number of concepts imply and-relations between the concepts and, thus, indicate complexes of findings. Arrows within a frame refer to increase or decrease (see page 72).

Appendix 3.3

Worked-out example of a comparison of a student's explanation to the canonical explanation of the case of heart failure

Worked-out example of the comparison of the student's explanation in Figure 3.1 to the canonical explanation in Appendix 3.2. The student applied 14 concepts in her explanation. These concepts and their causal links have been drawn onto the canonical explanation. The model concepts are shaded (concept no. 1, 2, 6, 7, 8, 11, 12, and 13), resulting in a total number of 7. The detail concepts used by the student are indicated with "detail" (concept no. 3, 4, 5, and 9), resulting in a total number of 4. The student used 1 model link (linking concept no. 6 and 7). The short cuts used by the student are represented by bold curved arrows (linking concept no 7 and 8, 11 and 12, and 11 and 13), resulting in a total number of 3. The alternative links used by the student are represented by curved arrows (linking concept no. 1 to 6, and 6 to 11), resulting in a total number of 8 (see page 73).





The explanation of clinical concepts by expert physicians, clerks and advanced students¹

The present study aimed to discover differences in the content and organization of medical knowledge in subjects with different clinical experience. Advanced students, clerks and experienced physicians were required to explain 20 current clinical concepts to the experimenter in approximately 2 minutes per concept. The subjects' explanations were analyzed on elaborateness, quality and fluency in which they were provided. The results showed that the more experienced subjects generally provided more elaborate, qualitatively better and more fluent explanations. This suggests that causal pathophysiological knowledge does not decay with clinical experience, but rather forms a coherent knowledge structure that can be easily accessed. The present findings, therefore, support the hypothesis that in the development of medical expertise networks of detailed pathophysiological knowledge relating causes and consequences of disease become encapsulated into clinically relevant concepts. Qualitative analysis, furthermore, showed that the differences between the three subject groups were not as large for all 20 concepts, suggesting that clinical experience is not the only mediator in forming effective encapsulations. Implications for medical education are discussed.

The development of medical expertise has often been studied by comparing the performance of medical students and expert physicians on tasks such as diagnosing, recalling and explaining clinical cases. Diagnostic competence certainly is the most direct measure of medical expertise. As might be ex-

¹Van de Wiel, M. W. J., Boshuizen, H. P. A., Schmidt, H. G., & Schaper, N. C.. The explanation of clinical concepts by expert physicians, clerks and advanced students. (manuscript under review). Parts of this chapter appeared as Van de Wiel, M. W. J., Schaper, N. C., Boshuizen, H. P. A., & Schmidt, H. G. (1995). *The explanation of encapsulating concepts by medical experts, clerks and advanced students*. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, April 18-22. (ERIC Document Reproduction Service No. ED385194)

pected, studies in this domain invariably found that diagnostic accuracy increased with level of expertise. It was also found that expert physicians often used very little time to provide a diagnosis (Elstein, Shulman & Sprafka, 1978; Norman, Rosenthal, Brooks, Allen & Muzzin, 1989; Van de Wiel, Boshuizen & Schmidt, in preparation). Not only were quantitative differences found, think-aloud studies of diagnostic reasoning also showed that expert physicians and students used qualitatively different knowledge in diagnosing a clinical case. Whereas experts predominantly used clinical disease knowledge, students applied their detailed pathophysiological knowledge of disease processes (Bordage & Lemieux, 1991; Boshuizen & Schmidt, 1992). Clinical case recall studies, furthermore, showed that experts' recall protocols were less elaborate but contained more high-level inferences than those of students of an intermediate level of expertise (Boshuizen, 1989; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993b; Van de Wiel, Boshuizen & Schmidt, 1994b). Experts, thus, appeared to represent the clinical cases at a more summarized level than students. Clinical case explanation studies, finally, demonstrated that expert physicians provided less extensive, although more coherent pathophysiological explanations of the signs and symptoms in a case than advanced students (Patel, Evans & Groen, 1989a; Schmidt & Boshuizen, 1993b; Van de Wiel, Boshuizen & Schmidt, 1994a). In contrast to students, experts used virtually no detailed biomedical knowledge in their explanations. However, when confronted with difficult clinical cases expert physicians also used more detailed knowledge of pathophysiological disease processes in explaining the case information (Norman, Trott, Brooks & Smith, 1994; Patel, Groen & Arocha, 1990).

To explain these phenomena Schmidt and Boshuizen (1992) introduced the idea of *knowledge encapsulation*. Knowledge encapsulation is a form of knowledge restructuring in associative memory in which lower-level, detailed concepts and their interrelations are subsumed under a smaller number of higher-level concepts with the same explanatory power. It is assumed that in the development from medical student to expert physician networks of detailed biomedical concepts become encapsulated into more comprehensive clinical concepts or diagnostic labels. Thus, complex pathophysiological processes relating causes and consequences of disease are captured in a few encapsulating concepts that are clinically relevant. Through the direct activation of these encapsulating concepts in case processing, expert physicians can quickly understand and characterize a patient's problem. For example, in a study on the diagnosis, recall and explanation of a case of acute bacterial endocarditis, it was found that experts frequently used high-level clinical concepts such as septic syndrome, thrombo-embolism, aorta insufficiency and decreased bodily resistance (Schmidt & Boshuizen, 1993b).

Furthermore, Schmidt and Boshuizen (1993a; 1993b) assume that the detailed encapsulated knowledge is not lost, but, whenever necessary, can be retrieved through *unfolding* of the encapsulated knowledge structure. This

was clearly demonstrated in a study in which experts generated considerably more knowledge than advanced students when they were requested to tell everything they knew about endocarditis (Schmidt & Boshuizen, 1993b). Furthermore, when the high-level clinical knowledge is not adequate for interpreting a clinical problem, detailed biomedical knowledge will be used for diagnosis. For example, Norman et al. (1994) showed that experienced nephrologists used more causal pathophysiological explanations, and also more time than residents in diagnosing complex nephrology problems. This resulted in considerably better diagnoses for the nephrologists. The encapsulated structure of medical knowledge, thus, enables physicians to apply their knowledge flexibly to the task at hand. In the hectic practice of the clinic it allows them to process patient information efficiently, while in the discussion of complex problems (or recent scientific developments) or in the explanation to patients and students it allows them to provide easily a coherent explanation at the desired level of detail. For example, when a 55-year old man complains about bleeding piles, a bloated and tense feeling in the stomach and swollen veins around the navel, internists will directly activate the encapsulating concept of portal hypertension. On the other hand, when internists have to explain portal hypertension to students they will elaborate on the diseases and pathophysiological mechanisms underlying an elevated pressure in the portal venous system and the signs and symptoms this may cause in patients.

Finally, Schmidt and Boshuizen (1992; 1993b) assume that the learning of encapsulating concepts by medical students is induced when these students are confronted with real patients in medical practice. The extensive knowledge base which students build up in the preclinical part of their training consists of in-depth knowledge about the normal and pathological functioning of the body, and to a small extent of clinical knowledge about epidemiology, clinical features, diagnosis and treatment of diseases. The greater part of this knowledge is abstract knowledge, i.e. it is not experienced in real practice but rather acquired from textbooks, lectures, videos and eventually skills training. For students, therefore, it is a big change when they start to participate as clerks in clinical practice and become directly involved in the examination and treatment of patients. In their clerkships, students experience the variability in which diseases manifest themselves, the diagnostic methods applied, and the practice of therapy. In addition, they observe the way in which experienced physicians solve medical problems and communicate with each other. The newly acquired knowledge needs to be integrated with the present mass of biomedical knowledge into a coherent entity. Particularly in this phase of medical education the restructuring of knowledge in the form of encapsulation is starting to take place. Repeated application of detailed biomedical and clinical knowledge in diagnosing and treating patients will result in the encapsulation of elaborate networks of detailed concepts into clinically more applicable concepts. Many years of practi-

cal experience, however, are needed for medical graduates to develop a highly integrated knowledge base that is easily accessible while practicing clinical medicine.

Though designed for a different purpose, so far only the 1993b study by Schmidt and Boshuizen yielded evidence that indeed a large number of causally related pathophysiological concepts were associated with a higher-level clinical concept in experts' knowledge structures. When given the opportunity to activate their knowledge of endocarditis for 3.5 min., expert physicians came up with many more concepts than medical students. The number of concepts students activated, in addition, increased with years of medical training. To many readers these findings may seem rather obvious: The more professional training and experience, the more domain-specific knowledge a person will have. However, it is also known that detailed knowledge that is not used decays in memory (e.g., Anderson, 1983). This knowledge is, therefore, harder to access, or even "forgotten". Objective knowledge tests have indeed shown a decline in the performance of medical practitioners after 10 to 15 years of experience (Day, Norcini, Webster, Viner & Chirico, 1988; Norcini & Dawson-Saunders, 1994; Norcini, Lipner, Benson & Webster, 1985; Pollemans, 1994; Van Leeuwen et al., 1995). However, the detailed biomedical knowledge that is subsumed under encapsulating concepts is assumed not to decay, because it is integrated in a meaningful way with clinical disease knowledge. Encapsulation of biomedical knowledge is supposed to be a direct result of the tuning of knowledge to the problems encountered in medical practice, and the causal relations between biomedical and clinical knowledge are assumed to make the structure of medical knowledge highly coherent and stable. In this respect this line of thinking does not accord with Patel and colleagues who suppose that biomedical and clinical knowledge are organized as two worlds apart (Patel et al., 1989a; Patel, Evans & Groen, 1989b; Patel & Kaufman, 1995).

The present experiment was designed to investigate more thoroughly the idea of knowledge encapsulation by focusing on the content and organization of the knowledge subsumed under a number of current clinical concepts in subjects with different clinical experience. Advanced students, clerks and expert physicians were stimulated to tell as much as they knew about 20 clinical concepts that are frequently applied in the daily practice of internal medicine. If, as a consequence of clinical experience, detailed biomedical knowledge becomes encapsulated into functional units of clinical knowledge, it can be expected that experienced physicians will provide more elaborate, qualitatively better and more fluent explanations for the experimental concepts than medical students. In addition, it is expected that clerks who have just finished their clerkship in internal medicine would do better than advanced students.

Method

Subjects

Subjects were two groups of 15 medical students of Maastricht University, and 12 internists working at the affiliated university hospital. Medical students at Maastricht University follow a six-year undergraduate program: The first four years consist of preclinical courses provided within a problem-based curriculum, and the last two years comprise clinical clerkships. Students selected for the present experiment were fourth-year students who had completed their preclinical courses, and fifth-year students who had just finished their clerkship in internal medicine². The physicians' experience in internal medicine ranged from 7 to 29 years ($M = 14.9$). The students received a small compensation for their participation.

Materials

The central clinical concepts used in patient discussions at three sub-departments of internal medicine (gastroenterology, nephrology and general) were inventoried. An internist classified these concepts at one of six levels of encapsulation: A symptom, a complex of symptoms referring to bodily function, a complex of symptoms referring to patho-anatomy, more generalized, higher order concepts referring to bodily function or patho-anatomy, a working diagnosis, or a final diagnosis.

The concepts that were selected for the present experiment concerned the dysfunction of heart, liver and kidneys and addressed the most important underlying pathophysiological mechanisms and the resulting signs and symptoms. Not only were higher order encapsulating concepts chosen, such as right-sided heart failure or liver insufficiency, but also some general symptoms, such as edema. In addition, three concepts that were marked as highly encapsulating and which were often applied by experts in diagnosing a case of acute bacterial endocarditis (Schmidt & Boshuizen, 1992) were included. Finally, four concepts used in a pilot experiment were added; two of these were presented as practice items. In all, this resulted in a list of 20 experimental items (see Table 4.3).

Procedure

Subjects were told that the experimenter was interested in the meaning they gave to a number of medical concepts. They were instructed to explain each concept in approximately two minutes to the experimenter, who should be

² In the curriculum at Maastricht University the clerkship in internal medicine is offered as one of the first three clerkships. The other two are the clerkships in surgery and family medicine. The duration of each clerkship is 12 weeks. Thus, all the fifth year students participating in the present experiment had 12 weeks of experience in internal medicine and 0, 12, or 24 weeks of other clinical experience.

regarded as a fourth-year medical student. An example clarified again for the subjects that the experimenter wanted to hear an explanation of the medical concept, and how it could be recognized in a patient. The time was set at two minutes to prevent subjects from providing only a brief definition of the concept. Subjects were reassured that it did not matter if they had no precise knowledge of all concepts, but were encouraged to say if they did not know a concept or were not sure of their explanation; in the latter cases they were stimulated to guess as best as they could. All students participating in the present experiment were familiar with explanation tasks, because they followed a problem-based program in which students communicate acquired knowledge in small groups.

In order to prevent prior activated knowledge having different effects on subjects, the presentation order of the first eight concepts was the same for all subjects; so was the order of the concepts associated with one organ system. The presentation order of the organ systems (heart, liver, kidneys) was balanced to prevent a disproportionate position effect. The explanations were audio-taped and transcribed.

Analysis

To measure differences in availability and accessibility of knowledge in the three expertise groups subjects' explanations were analyzed on elaborateness, quality and fluency. The elaborateness of explanations was expressed as the number of concepts applied. Concepts were defined as small meaningful units of information in an explanation. Therefore, concepts that were exactly repeated in a subjects' explanation were only counted once, and metacognitive remarks, such as "I don't exactly know how this mechanism works", were not included. The nature of concepts, in addition, was coded in five categories. Concepts were coded as a definition, a cause of disease, clinical knowledge, pathophysiological knowledge, or therapeutic knowledge. Two examples of definitions given for the experimental concept of ascites are "ascites is the accumulation of fluid within the peritoneal cavity" and "ascites is, in fact, a special form of edema". Concepts were coded as a cause if they clearly referred to a possible cause of disease, such as "as a consequence of liver cirrhosis" or "portal hypertension can be caused by alcohol abuse". Concepts were coded as clinical if they concerned the signs and symptoms associated with disease, the frequency of occurrence of diseases and their manifestations, the degree of seriousness, the influence of patient characteristics or contextual factors, and the methods of examination. Examples are "inflammation is characterized by five local signs", "swelling of the ankles", "the most common complaint", "there is a risk of dying", "especially in older patients", and "measuring of creatinine in the blood". Concepts were coded as pathophysiological if they concerned the pathophysiological mechanisms of disease. Examples are "increased permeability of the capillaries", "higher levels of toxins in the blood", "increased secretion

of phosphate", and "if the right side of the heart has to pump against a higher resistance". Concepts were coded as therapeutic if they concerned the treatment or management of disease. Examples are "you have to give the patient antibiotics", and "disease that you have to treat as soon as possible".

The quality of explanations was determined by comparing subjects' explanations with canonical explanations of the experimental concepts. The canonical explanations included a definition of the concept, the major causes and clinical consequences, and the essential pathophysiological mechanisms of disease. They were based on recent medical literature (e.g., Isselbacher et al., 1994) and checked by specialists in internal medicine. In Appendix 4.1 the canonical explanation of the concept of inflammation is provided. The number of concepts in a subject's explanation that matched the model concepts in the canonical explanation were counted. Furthermore, a record was made of the number of clinical concepts presented that a subject did not know at all. A third parameter referring to the quality of knowledge measured the number of times a subject expressed him- or herself imprecisely in an explanation. This included concepts at too general a level of description, as well as concepts for which it was not fully clear what they referred to. For example, in defining ascites some students used the more imprecise concept of "ascites is the accumulation of fluid in the stomach" rather than "ascites is the accumulation of fluid in the peritoneal cavity". Other examples of imprecise expressions are "there must be something causing an elevated pressure", "you will get a kind of backward failure", "that kind of thing", and "it has to do with aldosterone".

The fluency of explanations, finally, was measured by five parameters. First, the number of times students and internists unexpectedly changed the topic in an explanation was counted. Second, the number of thinking pauses in an explanation was counted. A third parameter was the number of times a subject stumbled in providing an explanation. This parameter included the repetition of words and phrases, and unfinished or interrupted phrases. Furthermore, the number of times was counted that a subject was thinking aloud in an explanation. Examples of thinking aloud are "I think the blood pressure will decrease" and "this results in less angiotensin in the, or more?, let me think ...". Finally, the number of remarks was counted that referred to a lack of knowledge. Examples of these remarks are "what happens with the blood pressure, I really don't know" and "I cannot tell you more about it right now".

The protocols of 10 subjects were scored twice. Intrarater agreement ranged from .50 for the number of times subjects unexpectedly changed the subject, to .98 for the number of thinking pauses. The average intrarater agreement for all 14 parameters was .88. The data were analyzed by means of MANOVA or ANOVA applying level of expertise as independent variable. Polynomial contrast analysis was used to test if linear trends were present in the data. The Student-Newman-Keuls test (significance level of .05) was

used to make post-hoc comparisons between the three expertise groups. In addition, a more qualitative analysis of the data was performed. Finally, the validity of the quality measurement used in the present experiment was tested by correlating the mean quality scores students obtained for their explanations with the mean scores they obtained on two or three recent progress tests used in the Maastricht curriculum. The progress test is an overall test of medical knowledge consisting of 15 subject categories and totaling about 250 items.

Results and discussion

Elaborateness of explanations

The elaborateness of explanations as measured by the number of different concepts applied in each explanation showed a significant main effect for level of expertise ($F(2, 39) = 8.68$, $MS_e = 113.2$, $p = .0008$), which could be specified as a linear trend ($F(1, 39) = 17.34$, $MS_e = 226.1$, $p = .0002$). Thus, the more experienced the subjects, the more elaborate their explanations. Pair-wise comparisons between the three expertise groups, furthermore, revealed that internists applied significantly more concepts in their explanations than clerks, who, in turn, applied significantly more concepts in their explanations than fourth-year students. These results indicate that more experienced subjects have indeed access to a larger knowledge base than subjects with less clinical experience.

The analyses of the kind of concepts used in the explanations, in addition, showed that experienced physicians not only have more knowledge about the clinical aspects of disease, but also about the pathophysiological mechanisms underlying diseases. Significant main effects of expertise level were found for the number of definitions used in the explanations ($F(2, 39) = 11.39$, $MS_e = .64$, $p = .0001$), the number of causes of disease ($F(2, 39) = 3.64$, $MS_e = 2.01$, $p = .036$), the number of clinical concepts ($F(2, 39) = 13.83$, $MS_e = 5.08$, $p = .011$), and the number of pathophysiological concepts ($F(2, 39) = 24.94$, $MS_e = 3.61$, $p = .037$). Polynomial contrast analysis revealed that these effects could be specified as linear trends: In each of these four categories subjects of higher levels of expertise applied more concepts in their explanations. Interestingly, post-hoc comparisons between the fourth-year students and the clerks showed that clerks used significantly more clinical concepts than fourth-year students, but not more pathophysiological concepts. This suggests that medical students during their clerkship in internal medicine are more strongly focused on the clinical aspects of disease than on the pathophysiological mechanisms underlying diseases. The number of therapeutic concepts in the explanations was not affected by expertise level ($F(2, 39) = .84$, $MS_e = .60$, $p = .44$), presumably, because subjects were not explicitly asked nor encouraged to come up with methods to treat or cure patients.

Overall the subjects used 50.3 % pathophysiological concepts, 32 % clinical concepts, 11% causes of disease, 3.4 % therapeutic concepts, and 3.2 % definitions. In Tables 4.1 and 4.2 an overview of the results obtained in this study is provided.

Quality of explanations

The quality of explanations was expressed as the number of concepts in a subject's explanation that matched the model concepts in a canonical explanation. The main effect of level of expertise on the number of model concepts was significant ($F(2, 39) = 18.63$, $MS_e = 20.86$, $p = .0001$). Polynomial contrast analysis revealed a linear trend in the data ($F(1, 39) = 35.97$, $MS_e = 40.29$, $p = .0001$): The quality of explanations increased with level of expertise. This indicates that the more experienced subjects have access to a more complete and qualitatively better knowledge base than subjects with less clinical experience.

Table 4.3 shows the mean quality scores of the explanations of the three expertise groups for each of the 20 concepts. Fourth-year students had an extremely low score ($M = 1.5$) for the concept of passage complaints. It appeared that 11 out of 15 students were completely unfamiliar with this concept. Passage complaints is a typical clinical (Dutch) concept used in gastroenterology, which is according to one of the internists "reserved for passage complaints of the esophagus, and not referring to complaints of swallowing, but the patient feels that the food does not go through the esophagus". Most students had never heard the concept before, and the closest guess was "Oh, I think it has to do with constipation in the large intestine", while most clerks (13 out of 15) had picked up the term during their clinical rotations. More generally, both students and clerks had the lowest scores for concepts related to the organ system of the kidneys. Because the kidneys are an important part of the preclinical curriculum, these results indicate that this is a difficult subject for them.

The number of clinical concepts presented that were completely unknown to a subject was analyzed by means of ANOVA revealing a significant main effect of level of expertise ($F(2, 39) = 8.58$, $MS_e = 5.57$, $p = .00008$). The effect could be specified as a linear trend ($F(1, 39) = 14.78$, $MS_e = 9.60$, $p = .0004$). Besides passage complaints, these concepts were hepatic encephalopathy (unknown to 1 fourth-year student), uremic complaints (unknown to 3 fourth-year students and 1 clerk), decreased creatinine clearance (unknown to 1 fourth-year student), disturbances in calcium and phosphate metabolism (unknown to 2 fourth-year student and 1 clerk), and hypervolemia based on renal failure (unknown to 1 fourth-year student).

Finally, analysis of the number of imprecise expressions in the explanations showed again a significant main effect of expertise level ($F(2, 39) = 7.84$, $MS_e = 2.54$, $p = .0014$). However, both a linear and a quadratic component were significant ($F(1, 39) = 4.79$, $MS_e = 1.55$, $p = .035$, and $F(1, 39) = 11.81$, MS_e

Table 4.1. Means for the parameters measured in the explanation protocols, and an overview of the results of post-hoc analysis.

	Students <i>N</i> = 15	Clerks <i>N</i> = 15	Internists <i>N</i> = 12
Elaborateness of explanations			
Total number of concepts	17.4 ^{c,i}	20.2 ^{s,i}	23.2 ^{s,c}
Defining	.50 ⁱ	.60 ⁱ	.93 ^{s,c}
Cause	1.8 ⁱ	2.3	2.6 ^s
Clinical	5.4 ^{c,i}	6.6 ^s	7.3 ^s
Pathophysiological	9.0 ⁱ	9.7	11.7 ^s
Therapy / management	.57	.94	.62
Quality of explanations			
Number of model concepts	5.4 ⁱ	6.1 ⁱ	7.9 ^{s,c}
Unknown concepts ^a	.87 ⁱ	.40	.00 ^s
Imprecise expressions	1.35 ⁱ	1.74 ⁱ	.87 ^{s,c}
Accessibility of explanations			
Changing subject	.24 ⁱ	.24 ⁱ	.04 ^{s,c}
Thinking pauses	3.1 ⁱ	2.6 ⁱ	.25 ^{s,c}
Stumbling	4.3 ⁱ	5.4 ⁱ	2.9 ^{s,c}
Thinking aloud	1.0 ⁱ	.79 ⁱ	.14 ^{s,c}
Remarks referring to a lack of knowledge	1.0 ⁱ	.53	.17 ^s

^a Analyzed by means of ANOVA; ^{s,c,i} In post-hoc analysis significantly different from students, clerks and internists, respectively (sign. level $p < .05$).

= 3.83, $p = .0014$, respectively). The average number of imprecise expressions was 1.35, 1.74 and .87 for fourth-year students, clerks and internists, respectively. Pairwise comparisons between the three expertise groups showed that the internists used significantly fewer imprecise expressions than both the fourth-year students and clerks, whereas the use of imprecise expressions was not significantly different for the two student groups. This indicates that the knowledge of advanced students and clerks is not so well-developed that they automatically use the accurate concepts at an adequate level of detail.

Fluency of explanations

The fluency in which an explanation is formulated reflects both the availability and accessibility of knowledge. Expertise level had a significant effect on the number of times someone unexpectedly changed the topic in an explanation ($F(2, 39) = 9.78$, $MS_e = .18$, $p = .0004$). Both a linear and a quadratic component were present in the data ($F(1, 39) = 15.21$, $MS_e = 2.8$, $p = .0004$,

Table 4.2. An overview of the results of MANOVA and linear contrast analysis (*p*-values).

	MANOVA			Linear contrast
	Expertise	Concepts	Expertise	
			x Concepts	
Elaborateness of explanations				
Total number of concepts	< .05	< .05	< .05	< .05
Defining	< .05	< .05		< .05
Cause	< .05	< .05	< .05	< .05
Clinical	< .05	< .05	< .05	< .05
Pathophysiological	< .05	< .05	< .05	< .05
Therapy / management		< .05	< .05	
Quality of explanations				
Number of model concepts	< .05	< .05	< .05	< .05
Unknown concepts ^a	< .05			< .05
Imprecise expressions	< .05	< .05		< .05
Accessibility of explanations				
Changing subject	< .05	< .05		< .05
Thinking pauses	< .05	< .05		< .05
Stumbling	< .05	< .05		< .05
Thinking aloud	< .05	< .05	< .05	< .05
Remarks referring to a lack of knowledge	< .05	< .05	< .05	< .05

^a Analyzed by means of ANOVA

and $F(1, 39) = 5.48$, $MS_e = .10$, $p = .025$, respectively). Pairwise comparisons between the three expertise groups showed that internists unexpectedly changed the topic in an explanation less often than both fourth-year students and clerks (see Table 4.1).

Analysis of the number of thinking pauses in the explanations showed a significant main effect of expertise level ($F(2, 39) = 6.85$, $MS_e = 29.71$, $p = .0028$), that could be specified as a linear trend ($F(1, 39) = 12.31$, $MS_e = 53.39$, $p = .0012$). Internists hardly paused in their explanations, while clerks and fourth-year students used on average 2.6 and 3.1 thinking pauses in an explanation, respectively. Student-Newman Keuls tests were significant for the pairwise comparisons of internists with fourth-year students and clerks. Students paused when they did not know how to continue and were then obviously searching for the required knowledge.

Expertise level had also a significant effect on the number of times subjects stumbled in explaining a concept ($F(2, 39) = 7.53$, $MS_e = 20.20$, $p = .0017$). Both a linear and a quadratic component were present in the data ($F(1, 39) =$

Table 4.3. Mean quality scores for the explanations of each experimental concept by fourth-year students, clerks and internists.

Concepts	Students N = 15	Clerks N = 15	Internists N = 12
Inflammation (24)	8.4	8.2	9.8
Thrombo-embolism (30)	7.1	7.9	11.4
Edema (34)	7.5	8.1	8.0
Decreased bodily resistance (29)	5.5	6.1	6.7
Passage complaints (16)	1.5	4.8	6.4
Ascites (26)	4.6	6.3	8.2
Hyperventilation (47)	6.6	6.9	8.5
Sepsis (39)	7.5	7.3	9.2
Forward failure (39)	4.7	5.3	6.2
Pulmonary congestion (24)	5.9	5.7	5.6
Right-sided heart failure (20)	5.4	5.8	6.9
Respiratory acidosis (22)	4.8	5.1	7.0
Hepatic encephalopathy (34)	5.4	6.0	8.8
Portal hypertension (32)	9.5	8.0	9.8
Hepatic insufficiency (33)	5.8	5.9	10.4
Uremic complaints (43)	3.3	5.0	8.3
Decreased creatinine clearance (22)	3.3	4.5	6.5
Disturbances in calcium and phosphate metabolism (21)	3.8	6.7	7.4
Hypervolemia based on renal failure (26)	3.1	4.3	6.0
Thrombocytopathy (13)	4.7	4.6	6.6

Note—Quality scores were expressed as the number of model concepts in a subject's explanation. The total number of model concepts in the canonical explanations are provided between brackets behind each experimental concept.

4.70, $MS_e = 12.60$, $p = .036$, and $F(1, 39) = 11.26$, $MS_e = 30.22$, $p = .0018$, respectively). Pairwise comparisons showed that internists stumbled significantly less than both fourth-year students and clerks, who did not significantly differ from each other (see Table 4.1).

Furthermore, analysis of the number of times subjects were thinking aloud in an explanation showed a significant main effect of level of expertise ($F(2, 39) = 14.88$, $MS_e = 2.72$, $p = .0001$), that could be specified as a linear trend ($F(1, 39) = 28.52$, $MS_e = 5.22$, $p = .0001$): The more experienced the subjects, the less they were thinking aloud in providing their explanations.

Finally, analysis of the number of remarks referring to a lack of knowledge revealed a significant main effect of level of expertise ($F(2, 39) = 10.08$, $MS_e = 2.35$, $p = .0003$), that again could be specified as a linear trend ($F(1, 39) = 19.73$, $MS_e = 4.59$, $p = .0001$). Internists made just a few remarks that they were not sure because a concept was outside their specialty area ($M = .17$); occasionally clerks showed that their knowledge fell short ($M = .53$), while fourth-year students were frequently confronted with the limits of their knowledge in explaining the experimental concepts ($M = 1.0$).

These results clearly showed that internists explained the experimental concepts much more fluently than both fourth-year students and clerks. Clerks, in turn, appeared to be more fluent than fourth-year students. In addition, the results suggest that especially fourth-year students do not yet have enough knowledge available to provide coherent and complete explanations for the experimental concepts, and often have difficulties in accessing the required available knowledge.

Qualitative analysis of the results

Table 4.2 shows that not only a significant expertise effect has been found on the elaborateness, quality, and fluency of explanations, but also a significant concept effect, indicating that subjects did not provide equally elaborate, good and fluent explanations for each concept. In addition, it has been found that the interaction effect between expertise level and concepts was significant for a considerable number of parameters. Thus, although our hypothesis that more experienced subjects provide better, more elaborate and more fluent explanations was generally confirmed, this does not seem to be true for each of the 20 concepts. We expected that the direct interaction with patients in the clerkship period would stimulate students to build up a coherent body of knowledge in which clinical signs and symptoms of disease are embedded within pathophysiological disease knowledge. We also expected that they would become familiar with the encapsulating clinical concepts that are used by experts in discussing their patients. However, in reviewing the quality of the explanations we found that only for six concepts did clerks provide clearly better explanations than fourth-year students. These concepts were passage complaints, ascites, uremic complaints, decreased creatinine clearance, disturbances in calcium and phosphate metabolism, and hypervolemia. As already noted the concept of passage complaints was completely unknown to most of the fourth-year students. They often said so and frequently thought aloud. Although clerks provided much better explanations, they often stumbled in explaining the concept. This was also found in clerks' explanations of the concept of ascites. The concepts related to the functioning of the kidneys seemed, in general, difficult for both students and clerks: Not only did they provide explanations of a relatively low quality, also the number of times they paused, thought aloud and referred to their lack of knowledge was high. However, clerks

provided more elaborate explanations containing more clinical concepts and more model concepts than fourth-year students, indicating that they had learned about kidney diseases in the clinic. For example, with regard to the concept of uremic complaints most students and clerks knew that uremia is defined as a higher level of urea in the blood, but many students did not know any of the associated complaints, whereas most clerks knew some. These results show, that during the clerkship period medical students learn high-level clinical concepts, and improve their knowledge of the causes and consequences of disease. However, they still need more practice before their knowledge is as accurate and accessible as that of experts.

For six other concepts, we found that the knowledge of these concepts did not improve in the clerkship period, but later in medical practice. These concepts were thrombo-embolism, respiratory acidosis, hepatic encephalopathy, hepatic insufficiency, thrombocytopenia and sepsis. Internists provided much better and more elaborate explanations for these concepts than clerks and students. Clerks did provide more elaborate explanations than fourth-year students, and used more clinical concepts. In fact, clerks even used as many clinical concepts as internists in explaining most of these concepts (not sepsis). However, their explanations did not accord better with the canonical explanations than those of fourth-year students. In addition, both students and clerks were very imprecise in expressing themselves and used many thinking pauses. Students also made many remarks referring to a lack of knowledge. Thus, although clerks acquired some knowledge about these concepts during the clerkship of internal medicine, they had, presumably, not so often been confronted with these kind of clinical problems that this has led to coherent units of knowledge.

The explanations of four other concepts showed that effective encapsulation of knowledge can already take place before students have experience in medical practice. These concepts were pulmonary congestion, edema, decreased bodily resistance and portal hypertension. Both fourth-year students and clerks provided as good explanations for these concepts as internists. In addition, in all three expertise groups subjects hardly referred to a lack of knowledge, which clearly suggests that they felt they had mastered the subject matter. Most subjects explained the main disease process with the most important causes and consequences, and only differed in the details they specified. For example, in explaining pulmonary congestion most subjects mentioned left-sided heart failure as a cause of the accumulation of fluid in the lungs, and explained that leakage of fluid from the lung capillaries leads to pulmonary edema and shortness of breath. Thus, apparently, students and clerks had already integrated clinical aspects of disease with the pathophysiological mechanisms underlying these clinical problems. Interestingly, students and clerks even applied more clinical concepts in their explanations of edema than internists. However, also in their explanations of these four concepts, students and clerks used many imprecise expressions and fre-

quently stumbled. This suggests that fluent access to encapsulated knowledge is indeed the result of repeated application of knowledge in a clinical setting.

Furthermore, we expected that if experts have encapsulated coherent units of clinical and pathophysiological knowledge under higher level clinical concepts, internists would not only easily provide elaborate and high-quality explanations for these concepts, but would also build up their answer in a coherent way. This was indeed what we found. A typical internist's explanation had a clear beginning and end, addressed the definition of a concept, the causes and clinical consequences, and often ended with a short summary. This was also true for the good explanations of clerks, but most students' explanations were not that well organized. For example, in explaining the concept of hepatic insufficiency internists recapitulated the three functions of the liver: detoxification, synthesis and storage of metabolites, and worked out each of them. Students and clerks often only mentioned the function of detoxification. However, if they were later asked what they knew about the synthesis function of the liver, it appeared that most of them knew that the liver produces the blood clotting factors. As already noted, we also found that internists, in contrast to students and clerks, hardly ever unexpectedly changed the topic in their explanations.

Finally, the explanations in this experiment are a good source to trace important misconceptions in the understanding of the participating subjects. One of the misconceptions that frequently occurred concerned the dynamic relations between blood flow and blood pressure. Many students and clerks, and also some internists, were, for example, only talking about an accumulation of fluid behind the right or left ventricle, and did not mention the increasing blood pressure that developed in the supplying blood vessels. Interestingly, in explaining edema subjects usually mentioned that an elevated hydrostatic pressure caused the extravasation of fluid from the blood vessels to the intercellular space. Some other misconceptions concerned the concept of hyperventilation. Whereas some subjects correctly reasoned that a patient with hyperventilation has more oxygen and less carbon dioxide in the blood than usual because of a faster and deeper respiration, some other subjects reasoned that these patients have less oxygen in the blood because of a superficial respiration. There were also subjects who thought that the higher oxygen level instead of the lower carbon dioxide level was responsible for most of the clinical symptoms. For the purpose of medical teaching it would be valuable to systematically inventorize the misconceptions present in the explanations gathered in this experiment.

Validity of the quality measurement

To test the validity of the quality measurement in the present experiment the mean quality scores of the students' explanations were correlated with their scores on the Maastricht progress tests. Besides an overall score on the

Maastricht progress test (PT), separate scores are provided for items derived from the basic sciences, the clinical sciences and the social sciences. For each of the 30 students participating in the present study means were computed for these scores based on two or three recent progress tests. The mean scores on the progress test were correlated with the mean quality scores of the students' explanations. The correlation with the overall PT score was .40 ($p = .28$); the correlation with the basic sciences score was .22; the correlation with the clinical sciences score was .57 ($p = .0008$); and the correlation with the social sciences score was -.11. These results suggest that the knowledge subsumed under the clinical concepts explained in the present experiment is mainly based on knowledge from the clinical sciences, and to a much lesser degree on knowledge from the biomedical sciences. It seems, therefore, that the pathophysiological concepts used in the explanations are particularly meaningful from a clinical point of view.

General discussion

The results obtained in the present study clearly indicate that expert physicians can easily provide elaborate and high-quality explanations of current clinical concepts. Clinical experience, thus, entails rather than biomedical background knowledge is remembered as coherent units of information embedded within clinical disease knowledge than that it is forgotten as redundant knowledge. This supports the hypothesis that in the development of medical expertise biomedical knowledge becomes encapsulated into clinically relevant concepts, and that the encapsulated knowledge can be activated when this is required for task performance.

Further, it was expected that practical experience in the clinic would stimulate clerks to extend and restructure their knowledge resulting in a meaningful integration between biomedical and clinical knowledge. This was evidenced in the present experiment by the finding that the elaborateness, quality, and fluency of explanations increased with level of expertise. The clerkship period was only three months and for some concepts this clearly led to a better and fuller understanding of the subject matter. This was especially true for the concepts of passage complaints, ascites and uremic complaints. For some other concepts the explanations of clerks and students even equaled those of experts in quality (edema, pulmonary congestion and portal hypertension). However, clerks and students were less fluent, precise and coherent in formulating their explanations, suggesting that the available knowledge is not so easily accessible for them as for expert physicians. They often actively searched their knowledge base in constructing an explanation. In addition, the explanations of students and clerks also contained some serious misconceptions they were not aware of. It is obvious, therefore, that both fourth-year students and clerks still have to acquire

a great deal of knowledge and that they have to reorganize their knowledge into easily applicable encapsulating concepts which are used in medical practice. Clerks have already made a good start in their clinical rotations but this is by far not sufficient.

For medical education the present results imply that the clinical clerkships have a very important function in the integration of biomedical and clinical knowledge. Medical specialists supervising students in their clerkships need to realize that these students have not yet organized their knowledge around clinically relevant concepts, that they need to learn this in the clerkship period, and that they can be stimulated to do so. In the pre-clinical course of medical training, furthermore, the program needs to focus more on the relations between (patho)physiology, clinical methods of examination, and signs and symptoms of disease, so that students can learn the inherent relations between the concepts used in these domains. The best results will probably be obtained when this knowledge is grounded in systematic experiences with the examination of patients, so that experiential and conceptual knowledge are merged in one coherent body of knowledge. Finally, medical educators need to be alert to recognize and correct the misconceptions in their students' understanding of basic pathophysiological mechanisms of disease.

The explanation of important clinical concepts, that was used as the experimental method in the present study, appeared to be a valuable tool for students to assess their own knowledge. Students frequently reported that they needed to study certain subjects more precisely because they were confronted with the limits of their knowledge. In explaining the experimental concepts they clearly experienced to what extent they were familiar with the different aspects of a clinical problem; they not only discovered what aspects they did not know or were forgotten, but also which disease processes they did not really understand.

Research of Chi and colleagues (Chi, Bassok, Lewis, Reimann & Glaser, 1989; Chi, De Leeuw, Chiu & Lavancher, 1994), moreover, showed that the generation of self-explanations improves learning and understanding. Guided by accurate monitoring of their understanding "good" students spontaneously provided self-explanations while studying worked out examples of mechanics problems (Chi et al., 1989). Self-explanations also facilitated the integration of new information into existing knowledge when students were explicitly requested to self-explain in studying expository text on the human circulatory system (Chi et al., 1994). In medical education, therefore, explaining central clinical concepts to themselves and others could actually contribute to a better understanding of the underlying principles and resulting consequences of disease.

References

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Academic Medicine*, 66(9), S70-S72.
- Boshuizen, H. P. A. (1989). *De ontwikkeling van medische expertise: een cognitief-psychologische benadering* [The development of medical expertise; a cognitive psychological approach]. Doctoral dissertation. Haarlem: Thesis publishers.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-477.
- Day, S. C., Norcini, J. J., Webster, G. D., Viner, E. D., & Chirico, A. M. (1988). The effect of changes in the medical knowledge on examination performance at the time of recertification. *Proceedings of the Annual Conference on Research in Medical Education*, 27, 139-144.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving*. Cambridge, MA: Harvard University Press.
- Isselbacher, K. J., Braunwald, E., Wilson, J. D., Martin, J. B., Fauci, A. S., & Kasper, D. L. (Eds.). (1994). *Harrison's Principles of Internal Medicine* (13th ed.). New York: McGraw-Hill.
- Norcini, J., & Dawson-Saunders, B. (1994). Issues in recertification in North America. In D. Newble, B. Jolly, & R. Wakeford (Eds.), *The certification and recertification of doctors: Issues in the assessment of clinical competence*, (pp 28-46). Cambridge: Cambridge University Press.
- Norcini, J. J., Lipner, R. S., Benson, J. A., & Webster, G. D. (1985). An analysis of the knowledge base of practicing internists as measured by the 1980 recertification examination. *Annals of Internal Medicine*, 102, 385-389.
- Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125, 1063-1068.
- Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E. K. M. (1994). Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine*, 6(2), 114-120.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989a). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling*, (pp. 53-112). Cambridge, MA: The MIT Press.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989b). Reconciling basic science and clinical reasoning. *Teaching and Learning in Medicine*, 1(3), 116-121.
- Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition*, 18(4), 394-406.
- Patel, V. L., & Kaufman, D. R. (1995). Clinical reasoning and biomedical knowledge: Implications for teaching. In J. Higss & M. Jones (Eds.), *Clinical reasoning in the health professions*, (pp. 117-128). Oxford, UK: Butterworth Heinemann.
- Patel, V. L., & Medley-Mark, V. (1986). Relationship between representation of textual information and underlying problem representation in medicine. *Cognitive research series*. Montreal: McGill University, Centre for Medical Education. (CME86-CS1)
- Pollemans, M. (1994). *Kennistoetsing bij huisartsen* [Testing of knowledge in family physicians]. Doctoral dissertation. Maastricht: Universitaire Pers Maastricht.
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-282). New York, NY: Springer Verlag.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), 205-221.

- Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, 21(3), 338-351.
- Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (1994a). *Differences in knowledge application by students and medical experts in clinical reasoning*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, April 4-8. (ERIC Document Reproduction Service No. ED367709)
- Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (1994b). *The influence of lab data format and task instruction on clinical case recall*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, April 4-8. (ERIC Document Reproduction Service No. ED367710)
- Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (in preparation). Robustness of the intermediate effect in clinical case recall.
- Van Leeuwen, Y. D., Mol, S. S. L., Pollemans, M. C., Drop, M. J., Grol, R., & Van der Vleuten, C. P. M. (1995). Change in knowledge of general practitioners during their professional career. *Family Practice*, 12, 313-317.

Appendix 4.1

Canonical explanation of inflammation

(Boldly printed concepts are model concepts)

Inflammation is a reaction of the body (**inflammatory response**) to **tissue injury**.

It is clinically characterized by five **cardinal (local) signs**: the affected area is **red** (rubor), **swollen** (tumor), **hot** (calor), **painful** (dolor), the **function is disturbed** (functio laesa). **Systemic manifestations are fever** (mediated by interleukines) and increase in the numbers of circulating white blood cells (**leukocytosis**).

Causes of tissue injury:

Mechanical trauma, such as cutting or crushing.

Chemical injuries, such as those produced by acids, alkalis and phenols. An important cause is the presence of physiological substances in inappropriate locations.

Injury due to extremes of cold or heat (burns and frostbite).

Injury caused by living organisms such as bacteria, viruses, parasites, worms and fungi.

Ultraviolet or x-irradiation.

Internal injuries: injury due to a degree of reduction in the arterial blood supply sufficient to cause death of the underperfused tissue; injury produced by the inappropriate or excessive operation of immune mechanisms.

Inflammatory response:

Specific inflammatory mediators produced at the sites of injury regulate the response of the vasculature to injury.

Vasoactive mediators --> **vasodilatation** --> redness and warmth.

Increased blood flow and an increase in **permeability of the vessel walls** (loss of integration of endothelial cells):

--> **increased leakage of fluid** from the intravascular compartment into extravascular spaces
--> edema --> swelling;

--> **local stasis**.

Chemotactic factors are generated that recruit white blood cells (**recruitment of stimulation of inflammatory cells**) from the vascular compartment into the injured tissue.

Literature

Rubin, E. & Farber, J. (1990). *Essential Pathology*. Philadelphia, PA: Lippincott.

N. Woolf (1986). *Cell, tissue and disease: the basics of Pathology* (2nd ed.). London: Balliere Tindall.

A failure to reproduce the intermediate effect in clinical case recall¹

The present study is a replication of the 1993 recall study of Schmidt and Boshuizen with four different clinical cases from the domain of internal medicine. Subjects of four levels of expertise were required to study, diagnose and recall each case. The time available for case processing was manipulated. As expected, diagnostic accuracy increased with level of expertise. However, the study failed to disclose an intermediate effect in clinical case recall: A positive linear relation between expertise level and case recall was found. Constraining processing time did not affect diagnostic accuracy, but equally affected the recall performance of the various expertise levels. These results suggest that experts did not process case information in encapsulated mode, but represented the cases more elaborately while unfolding encapsulated knowledge.

Research on expertise in the past few decades has shown that expert performance is primarily a reflection of highly specialized domain knowledge (e.g., Ericsson & Smith, 1991). When experts solve a problem in their domain of expertise, they have rapid access to the relevant knowledge in long-term memory. Domain knowledge and information given in the problem interact resulting in a coherent representation of the problem at hand (Kintsch, 1988; Van Dijk & Kintsch, 1983). This rapid processing of information enables experts to produce a correct solution to the problem in a short time. Because experts have much more domain knowledge than less experienced persons, it is supposed that they will form more elaborate and more complete problem representations. Free recall of the information given in a problem has been frequently used to measure the elaborateness of the prob-

¹ Parts of this chapter appeared as Van de Wiel, M. W. J., Boshuizen, H. P. A., Schmidt, H. G. & P. W. de Leeuw. (1993). *A failure to replicate the intermediate effect in clinical case recall*. Paper presented at the Annual meeting of the American Educational Research Association. Atlanta, GA, April 12-17. (ERIC Document Reproduction Service No. ED362542)

lem representations. The rationale behind the use of this measure is that information which is encoded in problem solving will be retrieved in recalling the problem after it is solved. Thus, it is assumed that the more elaborate the problem representations, the more elaborate the recall will be. Studies on expertise in a great variety of domains, such as chess (De Groot, 1965; Chase & Simon, 1973), baseball (Spilich, Vesonder, Chiesi & Voss, 1979), bridge (Charness, 1979), electronic engineering (Egan & Schwartz, 1979), and computer-programming (McKeithen, Reitman, Rueter & Hirtle, 1981) indeed showed that experts remembered more information from a representative task in their domain of expertise than less experienced subjects. De Groot (1965), for example, presented chess positions for a very short time (2-10 seconds) to advanced chess players, a master and a grandmaster, and required them to report on their thoughts and perceptions during the brief presentation, and to recall the presented chess position as best they could. He found that the more expert chess players mentioned more often the best next move, and accurately recalled a larger number of the positions of chess pieces reaching a virtual perfect match of the chess position.

Studies on expertise in the domain of medicine, however, have generally not found that recall performance increases with level of expertise. The studies in this domain usually present a clinical case describing a patient's condition to subjects with different levels of medical training and practice, and request them to study, diagnose and recall the case. A large number of studies failed to find an effect of level of expertise on the amount of information recalled (e.g., Claessen & Boshuizen, 1985; Hassebrock, Johnson, Bullemer, Fox & Moller, 1993; Muzzin, Norman, Feightner & Tugwell, 1983; Muzzin, Norman, Jacoby, Feightner, Tugwell & Guyatt, 1982; Norman, Jacoby, Feightner & Campbell, 1979). Of those studies that found an effect of expertise on clinical case recall, only a few report a positive linear relationship of recall with level of expertise (e.g., Coughlin & Patel, 1987; Hobus, Schmidt, Boshuizen & Patel, 1987; Norman, Brooks & Allen, 1989), whereas a number of studies report an inverted U-shaped relationship of recall with level of expertise (e.g., Boshuizen, 1989; Patel & Frederiksen, 1984; Patel & Groen, 1991; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993). In an attempt to summarize the results of the recall studies in the medical domain, Boshuizen (1989) performed a meta-analysis on outcomes of 11 recall studies conducted between 1979 and 1987, which suggested that the inverted U-shaped relation better describes the combined results than a monotonically increasing function. Thus, in the medical domain, it was not experts but intermediate level subjects who recalled most information from the case material. This phenomenon was called the *intermediate effect* in clinical case representation studies (Schmidt, Boshuizen & Hobus, 1988).

Schmidt and Boshuizen (1992; 1993) explained this phenomenon of the intermediate effect by their theory of knowledge encapsulation. Knowledge encapsulation is a form of knowledge restructuring in associative memory

in which lower-level, detailed concepts and their interrelations are subsumed under a smaller number of higher-level concepts with the same explanatory power. Schmidt and Boshuizen supposed that through repeated application of knowledge in diagnosing and treating patients in medical practice, networks of detailed biomedical concepts become encapsulated into higher-level clinical concepts or diagnostic labels. They argued that if the knowledge of experts is organized in an encapsulated way, experts will activate higher level clinical concepts in case processing, and not detailed biomedical concepts, such as students do. Especially, intermediate level students who have acquired a large base of biomedical knowledge in the course of their medical training, but who have not yet a lot of clinical experience, will activate detailed biomedical knowledge in their attempt to comprehend a case. Experts' case representations, thus, will be more condensed than those of intermediate level students, and, hence, experts' case recall will be less elaborate.

On the basis of their theory of knowledge encapsulation, Schmidt and Boshuizen (1993) further expected that experts process cases at a faster rate than students, simply because they immediately activate a relatively small number of comprehensive, encapsulating concepts. These researchers, therefore, predicted that the case processing of experts would not be affected by a sizable reduction in processing time, whereas the elaborate processing of students would be impeded under these conditions. In fact, they predicted that the intermediate effect in clinical case recall would disappear when processing time was sufficiently reduced. Furthermore, they predicted that the case representations of experts and, hence, their recall protocols, would contain more high-level encapsulating concepts than those of students.

Two clinical case recall studies performed by Schmidt and Boshuizen (1993) confirmed these predictions and, thus, provided empirical support for their theoretical position. In the first study, they required health sciences students, second-, fourth- and sixth-year medical students and internists to study a case of acute bacterial endocarditis, to formulate a diagnosis, to recall the case information, and to provide a pathophysiological explanation of the signs and symptoms in the case. Processing time was limited to either 3 min 30 sec, 1 min 15 sec, or 30 sec. Diagnostic accuracy showed a positive linear relationship with level of expertise independent of processing time. As predicted, the relation between recall and expertise was different for each processing time condition: In the long processing time condition, second-, fourth- and sixth-year medical students provided more elaborate recall protocols than both health sciences students and internists, producing an intermediate effect in clinical case recall; in the 1 min 15 sec. condition no significant differences were found between subject groups; and in the extremely short processing time condition, a positive linear recall function with expertise was found. In all three processing time conditions, the more experienced subjects produced more summaries in their recall, a summary

being an inference based on two or more propositions from the text of a case.

In the second study, Schmidt and Boshuizen (1993) experimentally generated an intermediate effect in clinical case recall by applying a priming of prior knowledge paradigm. Health sciences students, second-, fourth- and sixth-year medical students and internists were requested to tell everything they knew about endocarditis for either 3.5 min or 30 sec. After completing the activation task all subjects were presented with the case of endocarditis used in the first study. They were requested to study the case for 30 sec, to provide a diagnosis and to recall the information in the case. Diagnostic accuracy showed a positive linear relationship with level of expertise independent of activation time. The amount of prior knowledge activated was higher in the long activation time condition. In addition, more prior knowledge was activated by subjects of higher levels of expertise. The recall data showed a clear inverted U-shaped function with level of expertise when subjects had the opportunity to activate their knowledge for 3.5 min. Moreover, second-, fourth- and sixth-year medical students remembered less from the case when they had less time to activate their knowledge. These results clearly suggest that the understanding of a clinical case by intermediates is dependent on the activation of detailed pathophysiological knowledge, whereas experts' understanding is independent of this kind of elaborate processing.

The effects of processing time in these two recall studies, thus, provide strong additional support for the idea that knowledge becomes encapsulated in the development from student to expert physician. Unfortunately, these studies were performed with only one clinical case, that of acute bacterial endocarditis. The same case of endocarditis, moreover, was used in two studies of Patel and colleagues that demonstrated an intermediate effect in clinical case recall (Patel & Groen, 1991; Patel & Medley-Mark, 1986). In order to broaden the base of studies documenting the generalizability and robustness of the phenomenon of the intermediate effect, it seemed, therefore, appropriate to replicate the 1993 study of Schmidt and Boshuizen, employing different case materials.

Subjects of four different levels of expertise were instructed to study clinical cases taken from four different specialty areas in internal medicine. Three processing time conditions allowed subjects to study the case for either 3 min, 1 min 15 sec, or 30 sec. They were required to formulate a diagnosis, and to subsequently write down what they remembered from the cases.

Method

Subjects

Subjects were 96 students and physicians of Maastricht University: 24 second-year, 24 fourth-year and 24 sixth-year medical students² and 24 internists with at least 4 years of experience in internal medicine. Each group of 24 was subdivided into three groups of eight who were assigned to three time constraint conditions. The subdivision was random for the three student groups, whereas it was balanced for the internists based on their subspecialisms. Subjects received a small compensation for their participation.

Materials

The materials consisted of four booklets, each containing a description of a clinical case and two blank response sheets. Each clinical case description reported some contextual information, the complaint, findings from history taking and physical examination, the relevant laboratory data and some additional findings. The four clinical cases presented were pheochromocytoma, stomach carcinoma, heart failure and liver cirrhosis. These were based on actual patients and were presented following a standard patient presentation format. Except for pheochromocytoma, these cases are fairly common in medical practice. The case descriptions were about half a page in length and consisted of 33, 42, 43, and 35 propositions respectively. The translated text of the case of stomach carcinoma is provided in Appendix 5.1.

Procedure

Subjects were told that four cases would be presented and that it would be their task to study a case carefully in order to produce a diagnosis and then to write down what they remembered from the case. Thus, for each case subjects subsequently studied, diagnosed and recalled the case. An example case was presented to familiarize them with the case format in this experiment and to experience the study time allowed. Depending on the experimental condition subjects were given the opportunity to study each case for 3 min, 1 min 15 sec, or 30 sec. Subjects were free to use as much time as they needed for the assignments. The order of case presentation was balanced.

Analysis

Diagnoses were scored on a scale ranging from 0 (completely inaccurate diagnosis) to 6 (completely accurate diagnosis) for each case. Points were

² The first and second years of medical education at Maastricht University emphasizes the acquisition of basic science knowledge, aiming at a good understanding of physiology, anatomy and general pathophysiology by students. In the third and fourth years education is centered around clinical problems addressing pathophysiological knowledge of disease, and the fifth and sixth years consist of different clerkships in the clinic.

awarded to accurate diagnostic elements. For example, the complete diagnosis for the stomach carcinoma case was "stomach carcinoma and atherosclerosis". Four points were given if the diagnosis contained the term "stomach carcinoma" and two points were given if it contained the term "atherosclerosis". Two points were given for partially correct diagnoses, such as "upper digestive tract carcinoma".

Based on a technique of proposition analysis for medical protocols (Patel & Groen, 1986), recall protocols were segmented into small, meaningful information units or propositions. These propositions consisted of two concepts connected by a qualifier, such as, *specification (spec)*, *location (loc)* or *time information (temp)*. For instance, the protocol fragment "(...) The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla (...)" consists of five propositions: 1. heart-spec (enlarged to all sides); 2. auscultation-spec (murmur); 3. murmur-spec (holosystolic); 4. murmur-loc (at the apex); 5. at the apex-spec (radiating towards the axilla). For each proposition in the free recall, it was decided whether it matched any proposition, or combination of propositions, in the stimulus material. The total recall score was determined by counting the number of literally recalled propositions, paraphrases, slightly imprecisely recalled propositions, and inferences matching one or more propositions in the case. The number of summaries, i.e. inferences referring to more than one proposition in the case, was counted separately as well. For example, the five propositions in the above recall fragment matched literally with the original text giving a recall score of 5, and a summary score of 0. However, the same information might be recalled as "The heart is enlarged to all sides and auscultation reveals mitral valve insufficiency. " In this latter recall fragment, the proposition "auscultation reveals mitral valve insufficiency" summarizes four propositions from the original text (auscultation reveals a holosystolic murmur at the apex radiating towards the axilla) giving a recall score of 2 and a summary score of 1.

The protocols were scored by the author and a research assistant. Interrater agreement of these procedures exceeded .98. Data were analyzed by means of repeated measures MANOVA with expertise level and processing time as between-subjects factors, and cases as within-subjects factor. Polynomial contrast analysis was used to test specific hypotheses.

Results

Diagnostic accuracy

Figure 5.1 shows diagnostic accuracy as a function of level of expertise and processing time. A significant main effect of level of expertise on diagnostic accuracy was found, $F(3,84) = 70.59$, $MS_e = 186.5$, $p = .0001$. Polynomial contrast analysis revealed only a significant linear trend ($F(1,84) = 211.6$, $MS_e =$

599.0, $p = .0001$): Thus, the more experienced the subjects, the better diagnoses they made. An increase in diagnostic accuracy is considered one of the most stable consequences of increasing medical expertise. Therefore, it can be concluded that the subjects' tasks in this experimental setting were ecologically valid. The effect of processing time was not significant, $F(3,84) = .52$, $MS_e = 1.39$, $p = .59$. Similar findings were reported by Schmidt and Boshuizen (1993). These findings suggest that the quality of the diagnosis is largely processing time-independent, at least within the limits of the processing times allowed in these studies. The analysis further showed a significant case effect ($F(3,84) = 49.42$, $MS_e = 108.7$, $p = .0001$) and a significant interaction between cases and level of expertise ($F(9,84) = 5.03$, $MS_e = 11.06$, $p = .0001$). Polynomial contrast analysis for each of the four cases revealed significant linear trends of diagnostic accuracy with level of expertise for all four cases. However, some of the cases seemed to be more difficult than others. Overall the lowest scores were obtained for the case of stomach carcinoma (2.0) and the highest scores were obtained for the case of liver cirrhosis (4.6) suggesting that the stomach carcinoma case was most difficult for the subjects and the liver cirrhosis case the easiest.

Recall

The relation between the average number of propositions recalled, level of expertise and processing time is depicted in Figure 5.2, showing a positive, linear recall function of expertise. Analyses revealed a strong overall effect of level of expertise ($F(3,84) = 10.27$, $MS_e = .39$, $p = .0001$), which could be specified as a highly significant linear trend ($F(1,84) = 29.90$, $MS_e = 1.15$, $p = .0001$), without a significant quadratic component ($F(1,84) = .88$, $MS_e = .034$, $p = .35$). This means that the higher the level of expertise, the more elaborate the subjects' recall was. Processing time had also a significant effect on recall ($F(3,84) = 59.76$, $MS_e = 2.29$, $p = .0001$), while no interaction effect of processing time and level of expertise occurred: All subjects provided more elaborate recall protocols under longer processing time conditions. These effects of expertise level and processing time on recall are even more convincing when we examine these effects for each separate case by means of ANOVA. It was found that both the main effect of expertise level, and the main effect of processing time were significant for each case. In addition, no significant interaction effects occurred. For each case, the data showed a positive linear relationship of recall with level of expertise with a significant linear component. With one exception, the quadratic component of recall with expertise level was not significant. The case of liver cirrhosis showed a significant quadratic component ($F(1,84) = 8.89$, $MS_e = 126.8$, $p = .0037$), which was due to a relatively low recall performance of the internists in the 1 min 15 sec processing time condition for this case.

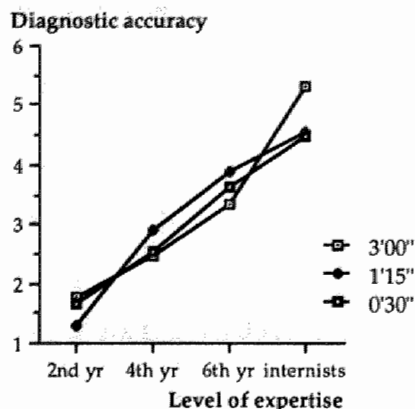


Figure 5.1. Diagnostic accuracy as a function of expertise level and processing time.

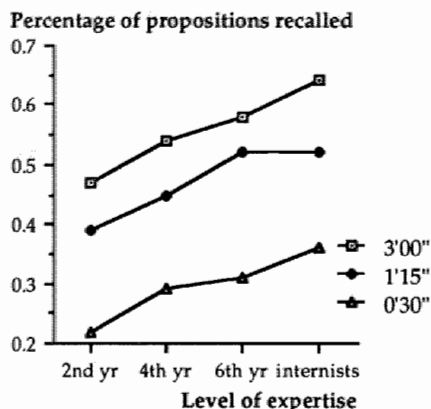


Figure 5.2. Percentage of propositions recalled as a function of expertise level and processing time.

Summaries

Figure 5.3 displays the average number of summaries produced in the recall protocols as a function of level of expertise and processing time. Summaries were defined by Schmidt and Boshuizen (1993) as inferences based on two or more propositions from the text. The main effect of level of expertise on number of summaries produced was significant ($F(3,84) = 27.88$, $MS_e = 10.62$, $p = .0001$); a linear component ($F(1,84) = 77.44$, $MS_e = 29.50$, $p = .0001$), as well as a quadratic component ($F(3,84) = 4.62$, $MS_e = 1.76$, $p = .035$) were present. In addition, pairwise comparisons between the four expertise groups by the Student-Newman-Keuls test (significance level of .05), revealed that sixth-year students produced more summaries than second- and fourth-year students, and internists, in turn, produced more summaries than sixth-year students. Processing time had no significant effect on the number of summaries produced, $F(3,84) = 2.75$, $MS_e = 1.05$, $p = .069$.

Discussion

The Schmidt and Boshuizen (1993) study showed that, under sufficiently long processing time conditions, medical experts remembered less from a case than advanced medical students. Their experts' recall was not affected by constraining processing time, whereas their students' recall was. In addition, they found that although experts' recall performance was lower than advanced students', the number of summaries produced by experts was greater. They assumed that these findings are the result of encapsulations in the knowledge base of experienced physicians. Because experienced physicians tend to process case information using knowledge in encapsulated

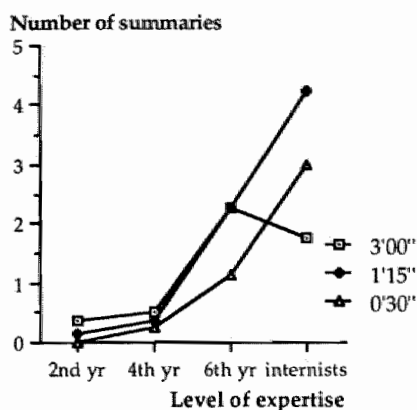


Figure 5.3. Number of summaries produced in recall as a function of expertise level and processing time.

mode, their recall was expected to be short and summarized rather than reproducing the information in the case. Students, on the other hand, did not yet have knowledge available in encapsulated format and, therefore, had to process the information in an elaborate fashion.

The present experiment largely failed to replicate these findings. An intermediate effect was not found. A comparison of the two data sets suggests that the discrepancy in recall performance between the two studies stems from a better recall performance of the experts and not from a poorer recall performance of the intermediate students in the present study: In the long processing time condition the average percentage of propositions recalled for fourth-year students, sixth-year students and internists was .55, .58 and .64, respectively, in the present study, while these outcomes were .56, .51 and .31, respectively, in the original study. These differences suggest that experts in the present study constructed more elaborate case representations than experts in the 1993 study.

With regard to the summaries produced in recall, the results compare more closely to those found in the previous study. In both studies, experts generally produced significantly more summaries than the student groups. Schmidt and Boshuizen (1993) interpreted these data as an indication that experts processed the case in encapsulated mode to a larger extent than students. We should note, however, that the average number of summaries produced in the present study is considerably smaller than in the study of Schmidt and Boshuizen: .77 versus 2.21, respectively.

The question, then, is: How can this failure to reproduce an intermediate effect in clinical case recall be explained? This question is particularly pertinent because the present findings are generally in agreement with those from other domains such as baseball (Spilich et al., 1979), bridge (Charness, 1979), electronic engineering (Egan & Schwartz, 1979), and com-

puter-programming (McKeithen et al., 1981). The latter were based on the assumption that experts construct more complete and coherent problem representations than students because they have developed a more elaborate knowledge base. On the other hand, intermediate effects do not represent an isolated phenomenon. In a large number of clinical case recall studies employing various experimental procedures and materials, it was found that experts provided less elaborate recall protocols than intermediate students (e.g., Boshuizen, 1989; Patel & Frederiksen, 1984; Patel, Groen & Frederiksen, 1986; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993). In addition, there is evidence from think-aloud studies that the granularity of the concepts used by experienced physicians while thinking about a case is indeed coarser than those applied by students (e.g., Boshuizen & Schmidt, 1992). This was also evidenced by post-hoc explanation studies which showed that experts in explaining the signs and symptoms in a clinical case applied more high-level explanatory concepts and less detailed biomedical concepts than students (Patel & Groen, 1986; Schmidt and Boshuizen, 1993; Van de Wiel, Boshuizen & Schmidt, in preparation). Furthermore, there is evidence from other domains that experts tend to chunk information in larger units for processing (e.g., Chase & Simon, 1973; Koedinger & Anderson, 1990; Rosenbloom, Laird, & Newell, 1988).

If we assume that the results of both studies are meaningful and the discrepancy in outcomes is not due to uncontrollable factors in the recall process itself, it is necessary to examine differences between both experiments in more detail.

First, the present experiment used four cases instead of one. Although subjects were instructed to study the cases in order to formulate a diagnosis, presentation of four cases in a sequential fashion may have led subjects to adapt to the requirements of the experiment and to pay special attention to information they would otherwise have processed in an encapsulated mode. If this is true, one would expect recall performance to improve during the course of the experiment, later cases being better recalled than earlier ones. Analysis of the effect of presentation order on the percentage of case propositions recalled showed indeed a significant effect, $F(3, 372) = 2.95$, $MS_e = .042$, $p = .033$. This position effect, however, cannot account for the failure to replicate the intermediate effect, because recall performance on the first case already showed a significant, positive, linear relationship with level of expertise: ANOVA on all cases which were presented first, revealed an overall effect of expertise level on percentage of propositions recalled ($F(3, 72) = 5.52$, $MS_e = .075$, $p = .0018$), which could be specified as a linear trend ($F(1, 72) = 15.40$, $MS_e = .21$, $p = .0002$), without a quadratic component ($F(1, 72) = .81$, $MS_e = .011$, $p = .37$).

Second, the four clinical cases used in the present experiment were different from the case used in the experiment of Schmidt and Boshuizen (1993). It might be possible that the four cases used in this experiment were

more difficult than the endocarditis case used by Schmidt and Boshuizen and, therefore, required elaborate processing of the data even by expert physicians. If this is true, it might be expected that experts needed more time to process the cases in this experiment, and diagnosed them less accurately. Informal observations indeed showed that most experts in the three-minute condition used all the time available to study the cases, whereas in the Schmidt and Boshuizen study, experts often after only one and a half minutes stopped processing, presumably because they did not need more time. The data on the quality of diagnoses, however, do not suggest that the cases in the present experiment were more difficult than the endocarditis case used by Schmidt and Boshuizen: Mean diagnostic accuracy of experts was 4.8 in the present study, and only 3.1 in the study of Schmidt and Boshuizen.

The present study consistently produced a monotone increasing recall function with expertise for four cases of internal medicine. This suggests that experts formed more elaborate case representations than students. Reduction of processing time, moreover, not only affected the recall performance of students but also that of experts: All subjects provided less elaborate recall protocols when they had shorter time to study the cases. This suggests that even the expert physicians processed the cases more elaborately when more study time was available. These findings are in contradiction with the findings of Schmidt and Boshuizen (1993) which suggested encapsulated processing of case information by experts. A comparison of the recall performance of experts in both studies indeed suggested that experts in the present study processed case information more elaborately than experts in the study by Schmidt and Boshuizen. This was also evidenced by the longer processing times of experts in the present study, and the lower number of summaries produced in their recall protocols. However, it remains unclear why the experts in the present study elaborately processed the cases.

References

- Boshuizen, H. P. A. (1989). *De ontwikkeling van medische expertise; een cognitief-psychologische benadering*. [The development of medical expertise; a cognitive psychological approach]. Doctoral dissertation. Haarlem: Thesis Publishers.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, 33 (1), 1-16.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education*, 19, 61-67.
- Coughlin, L. D., & Patel, V. J. (1987). Processing of critical information by doctors and medical students. *Journal of Medical Education*, 62, 818-828.
- De Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). The Hague: Mouton Publishers.
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, 7(2), 149-158.

- Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise. Prospects and limits*. New York, NY: Cambridge University Press.
- Hassebrock, F., Johnson, P. E., Bullmer, P., Fox, P. W., & Moller, J. H. (1993). When less is more: Representation and selective memory in expert problem solving. *American Journal of Psychology*, 106 (2), 155-189.
- Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). Contextual factors in the activation of first hypotheses: Expert-novice differences. *Medical Education*, 21, 471-476.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14, 511-550.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.
- Muzzin, L. J., Norman, G. R., Feightner, J. W., & Tugwell, P. (1983). Expertise in recall of clinical protocols in two specialty areas. *Proceedings of the 22nd Conference on Research in Medical Education*, 122-128.
- Muzzin, L. J., Norman, G. R., Jacoby, L. L., Feightner, J. W., Tugwell, P., & Guyatt, G. H. (1982). Manifestations of expertise in recall of clinical protocols. *Proceedings of the 21st Conference on Research in Medical Education*, 163-168.
- Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 1166-1174.
- Norman, G. R., Jacoby, L. L., Feightner, J. W., & Campbell, E. J. M. (1979). Clinical experience and the structure of memory. *Proceedings of the 18th Conference on Research in Medical Education*, 214-219.
- Patel, V. L., & Frederiksen, C. H. (1984). *Propositional representations of medical texts by medical students and physicians*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93-125). New York, NY: Cambridge University press.
- Patel, V. L., & Medley-Mark, V. (1986). Relationship between representation of textual information and underlying problem representation in medicine. *Cognitive research series*. Montreal: McGill University, Centre for Medical Education. (CME86-CS1)
- Rosenbloom, P. S., Laird, J. E., & Newell, A. (1988). The chunking of skill and knowledge. In H. Bouma and B. A. G. Elsendoorn (Eds.), *Working Models of Human Perception* (pp. 391-440). London: Academic Press.
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-282). New York: Springer Verlag.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, 21 (3), 338-351.
- Schmidt, H. G., Boshuizen, H. P. A., & Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 139-145.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 275-290.
- Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (submitted). Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Appendix 5.1

Case of stomach carcinoma

A 76-year-old male, who lives with his wife in a home for the elderly, is complaining of boring pain in his upper-abdomen for the last four months. His weight has decreased 9 kg in the last year. For this he is referred to a physician.

Further history taking reveals that he does not have complaints of nausea, but does have complaints of heart burn. The stool is sometimes a bit dark and he has some difficulty on swallowing. He gets easily tired. He has complaints of claudication with a walking distance limited to 750 m. He smokes 80 cigarettes a week.

Physical examination shows a healthy looking man. He weighs 52 kg and has a height of 1.58 m. The pulse is 84 per minute, and the blood pressure 170/80 mmHg. Auscultation reveals a loud aorta-sclerosis murmur. Over the abdominal aorta a murmur is heard, radiating to the iliac vessels. No abnormal palpable masses in the abdomen.

Laboratory results show an ESR of 41 mm/h (normal: < 7 mm/u) and a Hemoglobin-level of 8.2 mmol/l (normal: 8.5-11.0 mmol/l). Creatinine 84 μ mol/l (normal: 71-110 μ mol/l), ASAT (SGOT) 13 U/l (normal: < 40 U/l), ALAT (SGPT) 7 U/l (normal: < 40 U/l), LDH 354 U/l (normal: 200-450 U/l), alkaline phosphatase 77 U/l (normal: 30-125 U/l), gamma-GT 15 U/l (normal: < 50 U/l), total protein 57 g/l (normal: 65-97 g/l), albumin 32 g/l (normal: 35-55 g/l).

Ultrasound of the abdomen shows no abnormalities.

Knowledge encapsulation revisited:

Summary and general discussion

Early studies on the diagnostic reasoning of medical practitioners have clearly demonstrated that expert physicians generate a limited number of accurate diagnostic hypotheses very early in the patient encounter (Barrows, Norman, Neufeld & Feightner, 1981; Elstein, Shulman & Sprafka, 1978). A limited amount of patient related information forms the basis for these diagnostic hypotheses (Elstein et al., 1978; Hobus, 1994; Hobus, Schmidt, Boshuizen & Patel, 1987). If, for example, a 66-year-old man, who is known to be a heavy cigarette smoker, complains to his physician that he gets short of breath whenever he walks up the stairs to the first floor, the physician will tend to think of a heart problem. Prior knowledge of a patient's history, the patient's appearance and his nonverbal behavior, in addition to his presenting complaint, seem to interact in the thinking of physicians in the first minutes of an encounter. The diagnostic hypotheses that arise in the physicians' mind direct further questioning and examination of the patient. Initially, it was assumed that this process of diagnostic reasoning was characteristic of medical experts. However, the reasoning strategies employed by expert physicians and medical students were shown not to differ: Students also generated and tested diagnostic hypotheses while interacting with a patient. The diagnostic hypotheses of students, however, were less accurate than those of physicians (Elstein et al., 1978; Neufeld, Norman, Feightner & Barrows, 1981). These findings suggest that medical expertise is primarily based on a large and well-developed body of domain-specific knowledge. The study of medical expertise, therefore, would profit from an examination of the nature and organization of the knowledge structures underlying clinical problem solving, and their development during training and subsequent practice (e.g., Elstein et al., 1978; Feltovich & Barrows, 1984; Feltovich, Johnson, Moller & Swanson, 1984; Lesgold, 1984; Patel & Groen, 1986).

The development of medical knowledge structures in the course of medical training and clinical practice is the subject of the present thesis. Four studies

were performed to test a particular theory on the restructuring of medical knowledge proposed by Schmidt and Boshuizen, which is known as the theory of knowledge encapsulation (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992; Schmidt & Boshuizen, 1993b). In the next section, the theory of knowledge encapsulation will be recapitulated, as well as the empirical data that gave rise to it. Subsequently, the present research and its findings will be summarized and discussed. This will be followed by a section in which Schmidt and Boshuizen's theory on the development of medical expertise (1993a) will be reformulated by giving a more precise account of the changes in the knowledge structures in the growth from student to expert physician. Then, recommendations for future research will be provided. Finally, the implications of the present notions on the structure of medical knowledge for medical education will be elaborated upon.

Knowledge encapsulation

In the course of medical training students build up a large knowledge base on the normal and pathological functioning of the human body. They learn about the causes and consequences of disease processes and the factors that predispose the onset and development of disease in patients. They also learn about methods of history taking and physical examination, and treatment of disease. However, at the time students enter the clinic as clerks, the acquired knowledge is scarcely grounded in experiences with real patients. Only in the clerkship period, when they are directly involved in examining and treating patients, do they fully experience the variability in which diseases manifest themselves in patients. It is assumed that these practical experiences lead to the restructuring of students' knowledge bases, so that their knowledge becomes more apt to solving problems in clinical practice (Elstein et al., 1978; Feltovich & Barrows, 1984; Schmidt & Boshuizen, 1992; Schmidt & Boshuizen, 1993a). Years of clinical experience, however, are required for medical practitioners to tune their knowledge bases to the cases they encounter in their practice.

The knowledge that medical students and experts apply in clinical problem solving has been investigated in think-aloud studies of diagnostic reasoning (Boshuizen & Schmidt, 1992; Patel, Evans & Kaufman, 1990; Patel & Groen, 1986). In these studies, subjects of different levels of expertise are presented with a clinical case, usually consisting of a description of a patient's complaints and symptoms, the results of physical examination, the relevant laboratory data and the results of additional examinations. Subjects are required to think aloud while diagnosing the case. It was shown that medical students applied detailed knowledge of pathophysiological mechanisms of disease (Boshuizen & Schmidt, 1992; Patel, Evans & Kaufman, 1990). Medical experts, on the other hand, used hardly any detailed biomedical

knowledge in their diagnostic thinking, but predominantly used clinical disease knowledge (Boshuizen & Schmidt, 1992). These qualitative differences in the knowledge used by students and expert physicians in diagnosis eventually led Schmidt and Boshuizen (1992) to the idea of knowledge encapsulation. Knowledge encapsulation was defined as the subsumption, or packaging, of lower level detailed concepts and their interrelations in an associative net under a smaller number of higher level concepts with the same explanatory power. Schmidt and Boshuizen assumed that, as a consequence of repeated application of knowledge in diagnosing and treating patients in medical practice, elaborate networks of detailed biomedical concepts become encapsulated into more comprehensive clinical concepts or diagnostic labels. In this way, complex pathophysiological processes relating causes and consequences of disease are captured in a few encapsulating concepts that are clinically relevant. Through the direct activation of these encapsulating concepts in processing clinical information, experts can quickly understand and characterize a patient's problem. Students, on the other hand, have not yet developed knowledge in encapsulated form, and in interpreting clinical case information they will, therefore, elaborate on their detailed pathophysiological knowledge.

Support for the theory of knowledge encapsulation was provided in studies that investigated the mental representations students and physicians constructed in diagnosing clinical cases (e.g., Patel & Groen, 1986; Schmidt & Boshuizen, 1993b). In these studies, subjects of different levels of expertise were asked to study a clinical case, to provide a diagnosis, to recall the case, and/or to provide a pathophysiological explanation for the signs and symptoms in the case. Clinical case recall is assumed to reflect the content and structure of the representation formed in case processing. Pathophysiological explanation of a clinical case, in addition, is supposed to reflect the causal knowledge underlying the case representation. A frequent finding in these studies was that expert physicians represented a clinical case less elaborately than students of an intermediate level of expertise, a phenomenon known as the *intermediate effect* in clinical case representation (Schmidt, Boshuizen & Hobus, 1988). In recalling a clinical case expert physicians provided less elaborate protocols than intermediate students, but their protocols contained more high-level inferences that summarized various case elements (Boshuizen, 1989; Patel & Groen, 1991; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993b). In the study of Schmidt and Boshuizen (1993), for example, internists recalled a case of acute bacterial endocarditis using concepts such as septic syndrome, thrombo-embolism, aorta insufficiency and decreased bodily resistance. Students, on the other hand, recalled more precisely the information given in the case description, using concepts such as fever, chills, sweating, prostration and shortness of breath. The pathophysiological explanations that experts provided for the signs and symptoms in a case, furthermore, were more concise and of a higher quality than those of

intermediate students (Patel, Evans & Groen, 1989a; Schmidt & Boshuizen, 1993b): Experts used fewer concepts in their explanations, and these concepts were more closely related to a canonical case explanation. A canonical case explanation includes major lines of causal reasoning necessary to explain all the signs and symptoms in a case. These results suggest that experts represented the pathophysiology underlying a case at a more summarized or condensed level than students. In addition, it was found that internists, while explaining the case of acute bacterial endocarditis, applied the same high-level clinical concepts as in their recall of this case (Schmidt & Boshuizen, 1993b). Schmidt and Boshuizen accounted for this set of findings by formulating their ideas on the restructuring of knowledge in the development of medical expertise. They argued that the case representations of students were elaborate and detailed, because students reasoned through the extensive knowledge networks built up in medical training when attempting to comprehend a clinical case. The case representations of medical experts, on the other hand, were assumed to be condensed because experts could directly activate a limited set of relevant clinical concepts in an encapsulated form without having to resort to activating detailed biomedical knowledge. They further argued that if students rely on elaborate processing of a case, whereas experts use high-level encapsulating concepts in case processing, students, but not experts, would be impeded in case comprehension when processing time is constrained. This was indeed demonstrated in their 1993b study: Subjects of five levels of expertise studied, diagnosed, recalled and explained the case of acute bacterial endocarditis under three different time constraints. Recall and explanations of students were much less elaborate when less time was given for case processing. Recall and explanations of experts, however, were not affected by the restriction of processing time.

Knowledge encapsulation, as proposed by Schmidt and Boshuizen (1993a), finally, does not imply that detailed biomedical knowledge is lost in the growth towards clinical expertise. On the contrary, encapsulated knowledge about pathophysiological mechanisms of disease is assumed to be activated whenever necessary. In diagnosing *complex* clinical data, for example, expert physicians were found to make effective use of highly-specific pathophysiological domain knowledge (Norman, Trott, Brooks & Smith, 1994; Patel, Groen & Arocha, 1990). Furthermore, when subjects of various levels of expertise were explicitly asked to state as much as they knew about endocarditis, it was shown that experts generated considerably more knowledge than advanced students (Schmidt & Boshuizen, 1993b). The encapsulation of biomedical knowledge in expert physicians, thus, enables them to apply their knowledge flexibly to the task at hand.

Summary of the present studies

The influence of recall instructions and lab data format on clinical case recall

In Chapter 2 a large number of recall studies in the medical domain were reviewed. Most studies, in which a representative sample of novices, intermediates and experts participated, showed an intermediate effect in clinical case recall. However, there is one important exception: Norman, Brooks and Allen (1989) found that recall of clinical case material actually increased with level of expertise. It was argued that their findings may be dependent on the specific experimental materials they used in this study. They presented their subjects only with standard arrays of laboratory data. This contrasts with the usual practice in which subjects are presented with clinical cases consisting of a case history, physical examination data, relevant laboratory data and results of additional examinations. Norman and colleagues concluded from their findings that the interpretation of numerical lab data does not take place in an automatic, i.e. encapsulated fashion, but requires an effortful analysis even for expert physicians. In this chapter an alternative explanation for their findings was proposed, i.e. that the absence of relevant patient data, such as a case history and results of physical examination, led to effortful analytical processing of the lab data, because experts could not review these data in the light of relevant diagnostic hypotheses in the way they usually do in medical practice. Therefore, it was assumed that, if the lab data were presented in the context of a patient's history and physical examination, experts would generate diagnostic hypotheses rather automatically, enabling them to process the numerical lab data in encapsulated mode. Furthermore, it was assumed that under these conditions the differences in recall performance between a "memorize" and a "diagnose" instruction (i.e. an intentional vs. an incidental recall instruction) found in the Norman et al. study would disappear. If patient data were processed in an automatic or encapsulated mode, diagnostic hypotheses would be generated automatically, no matter what recall instructions were given.

To test these hypotheses a clinical case representation study was performed in which second- and fourth-year medical students, and internists studied, diagnosed and recalled three clinical cases under four different experimental conditions. The first clinical case in this study was presented under either incidental or intentional recall instructions. The other two clinical cases were presented in two versions: one with lab data in a numerical format and the other with lab data in an interpreted format. If experts could interpret the numerical lab data in encapsulated mode, as predicted, no differences in case representations were expected compared to the condition in which the numerical lab data had already been interpreted. On the other hand, if the interpretation of numerical lab data is an effortful analytical process, irrespective of the presence of other patient information, it was

expected that experts would represent case information more elaborately in the condition with numerical lab data.

The results supported the hypothesis that medical experts process clinical case information, including numerical lab data, largely automatically. No effects were found of recall instructions, nor of lab data format. Expertise level, however, affected study time, diagnostic accuracy and recall performance. Experts used less study time than both second- and fourth-year students, and provided much better diagnoses. Furthermore, experts' recall protocols were less elaborate than those of students, but contained more inferences that summarized case elements, again showing an intermediate effect in clinical case recall.

Pathophysiological representations of clinical cases

The study described in Chapter 3 aimed to investigate the restructuring of medical knowledge with the development of expertise by precisely comparing and contrasting the representation of pathophysiological knowledge underlying clinical cases in students and physicians. Second-, fourth- and sixth-year medical students and internists were instructed to study four clinical cases taken from different specialty areas in internal medicine. Three processing time conditions allowed subjects to study each case for either 3 min, 1 min 15 sec, or 30 sec. Subjects were required to provide a diagnosis and a pathophysiological explanation for the signs and symptoms in a case. Similar to the study of Schmidt and Boshuizen (1993b), the subjects' explanations were rewritten as semantic networks of meaningful concepts which were matched with canonical case explanations representing the major lines of causal reasoning necessary to explain all the signs and symptoms in a case. However, in contrast to the matching procedure used in the earlier study, not only the concepts matching a canonical explanation were considered, but also the concepts which were of a more detailed biomedical level of explanation. Moreover, the links between the concepts in subjects' explanations were considered, including the links matching those in a canonical explanation, the short cuts in reasoning as compared to the lines of reasoning in a canonical explanation, and the links connecting canonical model concepts in an alternative way than specified in a canonical explanation.

It was expected that with years of medical training students would provide better diagnoses, and produce more elaborate and detailed pathophysiological explanations that better matched the canonical explanations. However, as a consequence of the encapsulation of biomedical knowledge with clinical experience, it was expected that the elaborateness and level of detail of explanations would decrease beyond the intermediate level, whereas the quality of diagnoses and explanations would steadily increase.

The findings generally confirmed these predictions. Both diagnostic accuracy and the quality of explanations, as indicated by the concurrence of subjects' explanations with canonical explanations, increased with level of

expertise. The elaborateness and level of detail of the pathophysiological explanations, furthermore, indeed declined for experts as compared to advanced students. Experts used fewer concepts in their explanations, but the concepts they used were more confined to those specified in the canonical explanations. The same applied to the links experts specified between the concepts in their explanations. Experts also shortened the causal lines of reasoning that were specified in the canonical explanations more often than students by skipping intermediate concepts. Advanced students, on the other hand, used more detailed concepts than experts, and linked canonical model concepts more often in another way than was specified in the canonical explanations. These results clearly indicate that experts represent the pathophysiological knowledge underlying clinical cases at a highly summarized or condensed level, whereas students represent case information in a more elaborate and detailed way.

Constraining processing time, finally, was expected to affect case explanation in students, but not in experts. Students would be impeded in elaborating on their knowledge in studying the cases, and, therefore, provide less elaborate and accurate pathophysiological explanations. The results confirmed the hypotheses with regard to the quality of explanations, but not with regard to the elaborateness of explanations. Students provided qualitatively poorer explanations when they had only 30 seconds to study the cases, but this short processing time did not prevent them from providing elaborate case explanations. Experts, by contrast, provided explanations of the same high-quality in all three processing time conditions, but at a more summarized level of explanation when processing time was shorter. When experts had ample time to study the cases, they provided more elaborate explanations, addressing more canonical model concepts, as well as accounting in more detail for the underlying pathophysiological processes of disease. These findings suggest that experts are capable of processing case information in encapsulated form, but elaborate on their detailed encapsulated knowledge when given ample processing time.

This was not predicted by Schmidt and Boshuizen's theory of knowledge encapsulation (1992; 1993b). They assumed that the encapsulated networks of detailed biomedical knowledge form separate structures in the knowledge base that are only unpacked if required for task performance. Thus, in comprehending routine cases, these encapsulated structures would not be unpacked, and, hence, detailed biomedical concepts would not be activated. However, the present study showed that besides task difficulty, also processing time affects the activation of underlying pathophysiological knowledge. Therefore, Schmidt and Boshuizen's notions on the structure of expert knowledge were reinterpreted.

In the present view, medical knowledge is organized as a large associative network of concepts and their interrelations, encapsulations being simply the result of changing relations between these concepts. The present

findings, then, can be explained in terms of spreading activation through the knowledge network. Encapsulating concepts were assumed to be part of a coherent network of causal biomedical knowledge explaining a disease process. In these networks they are most strongly associated with the causes and consequences of the disease process. Therefore, in diagnosing a routine clinical case, encapsulating concepts will receive higher levels of activation than the associated detailed biomedical knowledge. However, if more time is spent processing the case information, all associated concepts in the knowledge network will receive higher levels of activation, and eventually become part of the case representation. This was indeed reflected in more detailed explanations by experts in the longer processing time conditions.

Explanation of encapsulating concepts

In Chapter 4 a study was presented that focused on the knowledge subsumed under encapsulating concepts. This knowledge is assumed to be organized in experts as stable and coherent networks of detailed pathophysiological concepts causally relating causes and consequences of disease. Although this detailed knowledge is not often used in clinical practice, it is not assumed to decay in course of time, because it is so meaningfully integrated with clinical disease knowledge. This contrasts with detailed biomedical knowledge that has been acquired in medical training but which is not related to the problems encountered in clinical practice (Norcini & Dawson-Saunders, 1994; Pollemans, 1994; Van Leeuwen et al., 1995). To gain more insight into the content and structure of the knowledge subsumed under encapsulating concepts in persons with different clinical experience, advanced students, clerks and internists were required to explain in detail 20 clinical concepts frequently used in the daily practice of internal medicine. For the explanation of each concept subjects had approximately two minutes. It was expected that if, as a consequence of clinical experience, coherent networks of detailed biomedical knowledge become encapsulated into functional units of clinical knowledge, expert physicians would provide more elaborate, qualitatively better and more fluent explanations for the experimental concepts than medical students. In addition, it was expected that clerks, who had only a 3-month experience in internal medicine, were so much stimulated to restructure and tune their knowledge on the basis of their interactions with patients that they would do better than advanced students, who had not yet had clinical training.

To test these hypotheses the subjects' explanations were compared on a large number of parameters. The findings were as follows: First, the elaborateness of explanations, as expressed by the number of different concepts used in the explanations, increased with level of expertise. The more experienced subject groups used more defining concepts, more causes of disease, and more clinical and pathophysiological concepts. Second, the quality of explanations was measured by the number of concepts in a subject's expla-

nation that matched the model concepts in a canonical explanation. It appeared that the more experienced subject groups indeed provided qualitatively better explanations. Six concepts were even completely unknown to some of the advanced students and clerks, and both of these subject groups used more imprecise expressions in their explanations than internists. Third, the fluency with which explanations were given increased with level of expertise: The more experienced subject groups used fewer thinking pauses, thought aloud less frequently, and referred less often to a lack of knowledge. In addition, internists stumbled less often in their explanations than students and clerks, and unexpectedly changed their train of thought less often. Finally, a more qualitative analysis of the results revealed that the differences in the quality of explanations between the three expertise groups were not as large for all 20 concepts. Thus, also without clinical experience relevant knowledge might be encapsulated into a clinical concept. However, in those cases, the underlying knowledge was not always that precise, usually still hard to access, and not yet very coherently organized. This suggests that encapsulation is not the result of a one-trial process, but rather of a gradual developmental process. Overall this study provides direct evidence that the knowledge subsumed under encapsulating concepts is not lost in the growth towards clinical expertise, but becomes more extensive and more coherently organized, so that it can be easily activated when it is required by the nature of the task.

Clinical case recall revisited

In Chapter 5, finally, a clinical case representation study was reported that showed an interesting anomaly. This study was set up as a replication of the 1993b recall study of Schmidt and Boshuizen. As stated previously, this study clearly showed an intermediate effect in the recall of a case of bacterial endocarditis, if subjects had ample time to study the case. However, if processing time was constrained, the intermediate effect disappeared. Schmidt and Boshuizen considered these findings important evidence for their theory of knowledge encapsulation: In contrast to students, experts did not process the case information in an elaborate form, but immediately interpreted it in an encapsulated form, resulting in more summarized case representations, and, hence, in less elaborate case recall. In the replication study reported in this thesis, second-, fourth-, and sixth- year medical students and internists, were presented with four clinical cases from the domain of internal medicine. Depending on the experimental condition subjects were instructed to study the cases for either 3 min, 1 min 15 sec, or 30 sec. For each case they were required to formulate a diagnosis and to recall the case information.

As might be expected, diagnostic accuracy was found to increase with level of expertise. However, in contrast to the findings of Schmidt and Boshuizen, no intermediate effect in clinical case recall was found, but a lin-

early increasing recall function with expertise. A comparison of the recall data in both studies revealed that the discrepancies in outcomes were due to a better recall performance of the experts in the replication study, and not to a poorer recall performance of the students. Constraining processing time, in addition, resulted in less elaborate recall not only in students, but also in experts. Finally, experts did produce more high-level inferences (summaries) than students in all three processing time conditions, but not to the same extent as experts in the study of Schmidt and Boshuizen. These findings clearly suggest that experts in the replication study processed and represented the case information more elaborately than experts in the original study.

Unfortunately, it was not possible to indicate any condition in the design and materials of these two studies that may have been responsible for the differences in outcomes between the two studies. The experimental task and the instructions given were similar. Although four cases instead of one were presented in the replication study, this did not affect recall performance. In addition, the cases used in the replication study were no more difficult than the case of bacterial endocarditis. Moreover, the recall study presented in Chapter 2 did show an intermediate effect in clinical case recall using three of the four cases presented in the replication study reported in Chapter 5. To gain more insight into the differences in performance of subjects in these three studies an overview of the results is provided in Table 6.1. As has already been argued, elaborate processing of cases costs time. Informal observations indeed suggested that experts in the study in Chapter 5 used more time for case processing than experts in the Schmidt and Boshuizen study. The present comparison clearly shows that they also used more time than the experts in the study in Chapter 2. Thus, presumably, they used this time to process the case information in a more elaborate way resulting in higher scores on diagnostic accuracy and higher recall performance.¹ Why they did so remains an unsolved question.

Interestingly, the method of free recall was used in the domain of medicine, because it was found to be a stable index of expertise in various other domains, such as chess (Chase & Simon, 1973; De Groot, 1965), baseball (Spilich, Vesonder, Chiesi & Voss, 1979), bridge (Charness, 1979), electronic engineering (Egan & Schwartz, 1979), and computer programming (McKeithen, Reitman, Rueter & Hirtle, 1981). In these domains, it was theorized that experts recall more information from a representative task in their domain than less experienced subjects, because the nature and organization of their knowledge allows them to perceive meaningful patterns of information that are encoded as chunks in memory. This seems not very different from the hypothesis that expert physicians interpret a clinical case in an

¹ The number of summaries produced in case recall seems related rather to the case material used than to the elaborateness of case processing.

Table 6.1. Overview of the results of three clinical case recall studies.

	Studies	2nd-year students	4th-year students	Internists
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Actual reading time	S & B (1993)	?	?	?
	Study Chapter 2	175 (12)	168 (32)	92 (34)
	Study Chapter 5	173 (13)	175 (14)	152 (33)
Diagnostic accuracy	S & B (1993)	.3 (.5)	1.0 (.8)	4.3 (1.3)
	Study Chapter 2	1.4 (1.4)	1.8 (1.8)	3.7 (1.7)
	Study Chapter 5	1.8 (1.9)	2.5 (2.1)	5.3 (1.4)
Percentage of propositions recalled	S & B (1993)	59 (18)	56 (13)	31 (14)
	Study Chapter 2	41 (12)	44 (12)	31 (17)
	Study Chapter 5	47 (11)	54 (12)	64 (18)
Number of summaries produced in recall	S & B (1993)	.25 (.46)	.13 (.35)	2.25 (1.75)
	Study Chapter 2	.02 (.14)	.19 (.39)	.67 (.95)
	Study Chapter 5	.09 (.30)	.13 (.34)	.44 (.62)

capsulated form. However, in encapsulation theory it is assumed that the encapsulating concept is reproduced in recall, and not the information that is summarized by the encapsulating concept. The clinical case recall study presented in Chapter 5, however, shows that this is not always the case. An interesting question, furthermore, is why experts in the other domains reproduced the information encoded in a chunk, and not the label representing that chunk in memory.

In part this may be due to the experimental procedures and materials applied in these studies. In most studies subjects were presented with a visual pattern of information for just a very short time, and subjects were asked to reconstruct the visual pattern as well as possible (Charness, 1979; Chase & Simon, 1973; De Groot, 1965; Egan & Schwartz, 1979). For example, when chess players are required to view a chess position for a few seconds, and to dictate the position from memory (De Groot, 1965) or to reconstruct this position by placing pieces on another board (Chase & Simon, 1973), it is fully clear why they recalled the actual positions of the chess pieces and not the labels for chunks of pieces.

Thus, there seem to be quite a few relevant differences between these recall studies and those performed in the medical domain. The most salient difference is that in the medical domain subjects engage in a problem solving task, recall being secondary to the formulation of a diagnosis. This was

clearly demonstrated by the study in Chapter 2, showing that, if subjects processed a case for diagnosis, recall performance was similar under intentional and incidental recall instructions. Second, in most clinical case recall studies the time given for case processing was ample or unlimited. However, if processing time was measured, experts were found to be faster than less experienced subjects (Boshuizen, 1989; Claessen & Boshuizen, 1985; Norman, Jacoby, Feightner & Campbell, 1979). Schmidt and Boshuizen (1993b) further demonstrated that, if processing time was sufficiently constrained, the fast and adequate case processing by experts was also reflected in higher recall performance. Third, although subjects in clinical case recall studies were asked to write down everything they remembered of a case, experts often seemed to confine their recall to a short summarizing description that characterized the patient's condition. However, the study in Chapter 5 showed that experts are, indeed, perfectly able to recall precisely the information in a case, recalling more case information than students. Fourth, within the medical domain there is the problem of case specificity (e.g., Elstein et al., 1978; Schmidt, Norman & Boshuizen, 1990): Not all cases may be represented at encapsulated level, either because a case is too atypical, or simply because relevant encapsulations may not be available in certain subdomains of medicine. These considerations cast some doubt on the usefulness of the method of free recall in expertise research, especially in the domain of medicine. In any case, it seems important to take these considerations seriously into account in designing and interpreting studies of free recall.

Reformulation of Schmidt and Boshuizen's theory on expertise development in medicine

In Schmidt and Boshuizen's theory on the development of medical expertise (1993a), knowledge encapsulation is presented as one of the cognitive mechanisms responsible for the transformation of intermediate into expert knowledge structures. Their theory describes expertise development as the progression through a sequence of distinct phases in which qualitatively different knowledge structures are used. It is assumed that in the first phase students build-up elaborate causal networks explaining the causes and consequences of disease in terms of underlying pathophysiological processes. Through the repeated use of knowledge in diagnosing and treating patients in medical practice, elaborate networks of causal knowledge would be encapsulated into high level diagnostic labels. At the same time knowledge would be tuned to the context and variability in which diseases manifest themselves in patients. Both processes, knowledge encapsulation and tuning, are considered to be responsible for the transition of knowledge from causal networks into narrative structures called illness scripts. In addition to the encapsulated pathophysiological knowledge of a disease, an illness script

is assumed to contain knowledge about patient background factors, complaints, signs, symptoms, diagnostic procedures, and course and treatment of the disease. In diagnosing a routine clinical case expert physicians would directly select and instantiate the appropriate illness script. The specific instantiations of scripts would remain available in memory as episodic traces of previously diagnosed patients, characterizing the final phase in expertise development. It is assumed that the knowledge structures acquired during the different phases of development form multiple layers in memory. Only when more recently acquired knowledge structures fail to produce an adequate representation of a clinical problem, knowledge structures acquired earlier would be accessed.

The major points of critique on the theory of Schmidt and Boshuizen (1993a) concern their assumptions on the discreteness of the phases in expertise development, and the unspecific ways in which the learning mechanisms and the illness script structure are described. The claim that the use of qualitatively different knowledge by medical students and expert physicians in clinical reasoning mark discrete changes in the structure of knowledge in acquiring expertise seems too strong, especially, when it is not clearly explained how a causal network evolves into multiple illness scripts. The differences in performance between subjects at various levels of expertise may characterize the preferred levels of representation at distinct stages of expertise rather than reliably reflecting the underlying developmental phenomena. In addition, easy access to knowledge structures acquired earlier implies that strong connections must exist between the knowledge in the multiple layers of memory. However, it is not clear how these connections are realized and developed. Finally, Schmidt and Boshuizen's stage theory does not enough take into account the case-specificity of medical expertise. Whereas some cases may be seen at least once a week, others are seen only twice a year, or even once a lifetime. Thus, the period required to develop a reasonable set of adequate illness scripts seems much longer than allowed in a stage theory of medical expertise. Moreover, the theory ignores that students, patients and laymen too may have developed illness scripts for the diseases they are frequently confronted with in their environment. Altogether, it seems that Schmidt and Boshuizen pushed too far the finding that students mainly rely on causal biomedical knowledge, whereas experts mainly rely on associative clinical knowledge represented in illness scripts.

The present proposal describes a more parsimonious theory on the restructuring of medical knowledge, assuming that expertise development is a rather gradual and continuous process. It is not assumed that the knowledge structures used by medical students and expert physicians are essentially different, but it is proposed that in the course of medical training and practice the knowledge network becomes larger, more accurate, and more coherently and adequately organized. The following section outlines how these changes are conceptualized in the knowledge network as students develop

into expert physicians. Subsequently, an illustration is given of how a beginning student, an intermediate student and a medical expert would use their knowledge in solving a routine clinical problem. Finally, the discussion addresses how experts would use their knowledge in diagnosing difficult cases, and in explaining clinical concepts to others.

The development of the knowledge network from student to expert

Like many other theorists (e.g., Anderson, 1983; Kintsch, 1988; Quillian, 1968) we assume that declarative knowledge is organized in an associative network of nodes and their interrelations. Nodes represent memory units² and links between nodes specify the strengths of the associations between these units. Association strengths are either positive or negative. The memory units dealt with in this thesis were confined to concepts or propositions that are meaningful in medical reasoning³. If a concept is activated in memory, activation spreads to concepts with which it is associated. The extent of spreading activation is proportional to the strength of the links between these concepts and the initially activated concept. Positive links have an excitatory effect leading to higher levels of activation, whereas negative links have an inhibitory effect. The spread of activation proceeds from node to node through the knowledge network. However, activation decays over time, so that at any moment only a tiny part of the knowledge network is activated.

Before entering medical school students have common world knowledge about illness and disease. This knowledge is usual associative or causal in nature, and refers to simplified, sometimes erroneous models of disease. When a person has fever, feels tired and has muscle pain, he will probably infer that he has the flu. In that case he also knows that it will take some days of bedrest to recover and that paracetamol can relieve the fever and the pain. In essence, this model could be considered as a predecessor of full-fledged illness scripts as described by Schmidt and Boshuizen (1993a). In a network theory the model can be described as a network with strong links between the nodes representing the symptoms of the patient and the concept of flu, as well as between the concept of flu and the concepts representing the beneficial effects of paracetamol (see also Yekovich & Walker, 1987). During the course of their medical training students build upon these models modifying the erroneous parts, adding associated findings, and incorporating causal chains of pathophysiological knowledge. For processes and diseases they are not familiar with, new models are developed. As students better understand the subject matter, their knowledge becomes more coherently organized. In medical practice, furthermore, knowledge becomes more and more tuned towards the cases encountered. Learning never stops

² Memory units represent either conceptual, propositional or perceptual knowledge.

³ For the purposes of this thesis we will treat concepts and propositions as similar entities.

in the medical profession but persists as long as physicians are involved in patient care and keep up with the specialist literature. Thus, in all stages of development, the acquisition of medical knowledge involves elements of accretion, restructuring, and tuning of knowledge (Rumelhart & Norman, 1978). In the earlier stages emphasis is on accretion and restructuring, whereas in later stages tuning becomes more important. The learning mechanisms responsible for the corresponding changes in the knowledge network are the addition of concepts and links, and the weakening or strengthening of links between concepts.⁴

Encapsulation of knowledge refers to the process that a high level clinical concept, describing a pathophysiological state, syndrome or disease process, becomes more strongly linked to concepts representing a specific set of signs and symptoms, as well as to concepts representing the possible causes of disease. The detailed network of concepts that relate the causes and consequences of disease by explaining the underlying pathophysiological mechanisms, therefore, becomes, as it were, by-passed. This network of pathophysiological knowledge itself is also coherently organized by strong causal links between concepts. However, the association strengths between encapsulating concepts, signs and symptoms and causes of disease are much stronger than the association strengths with and among more detailed pathophysiological concepts.⁵ Thus, encapsulating concepts acquire, in fact, a central status in these coherent networks of biomedical and clinical knowledge. Therefore, the knowledge "subsumed under an encapsulating concept" is nothing else than a specific subset of highly interrelated concepts in the knowledge network, that can be summarized by the encapsulating concept. This network structure of encapsulation can account for the finding in Chapter 4 that depending on their experience subjects explain high level clinical concepts more or less fluently, coherently and completely. If subjects have had the right training and experience, the knowledge associated with these concepts may have acquired "the status" of encapsulation, and, hence, these concepts may be fluently and adequately explained. On the other hand, if subjects' knowledge networks are not yet adequately organized, then, their explanations will be less fluent, coherent and complete. Thus, encapsulation is determined by the quality and coherence of the knowledge associated with the encapsulating concept.

⁴ In his ACT-R theory Anderson (1993; Anderson, 1996) argues that learning of declarative knowledge also involves the learning of base-level activations of nodes in the network. The base-level activation of a node reflects the usefulness of that node in the past. The activation level of a node in ACT-R is determined by base level activation and spread (contextual) activation.

⁵ The mechanisms mediating the strengthening and weakening of links between concepts may assumed to be based on a weighting process of the activation levels of concepts in diagnostic reasoning (Hebb, 1949; Marre, 1991), or on a Bayesian statistical estimation process reflecting the pattern of past use (Anderson, 1993; Anderson, 1996).

The formation of illness scripts is, arguably, similarly organized. However, whereas encapsulations mainly refer to intermediate diagnostic concepts, illness scripts rather seem to refer to the final diagnoses or causes of disease. In fact, illness scripts may be one step further than encapsulations, in the sense that they may also by-pass the encapsulating concepts. In routine cases, tuning of knowledge to its practical use by strengthening and weakening associations between concepts, may result in strong direct links between specific sets of patient background factors, signs and symptoms, causes of disease and ways of treatment. However, in many cases "illness script concepts" will also be strongly linked to encapsulating concepts. There is one clear exception: If no pathophysiological account of the disease is available, and, hence, encapsulations do not exist, only illness scripts representing associative knowledge can be developed. This is often the case in laymen, but, if scientific explanations for phenomena are still lacking, also in experts. Interestingly, Custers (Custers, 1995; Custers, Regehr & Norman, 1996) who studied extensively the development and function of illness scripts as proposed by Schmidt and Boshuizen (Boshuizen, 1989; Schmidt & Boshuizen, 1993a; Schmidt et al., 1990), eventually suggested that illness scripts might be organized in an associative network, as well.

Diagnosis of routine clinical cases by students and experts

In diagnosing a clinical case, incoming information activates corresponding concepts in the knowledge network. Activation spreads through the knowledge network and only those concepts that are highly activated will be consciously represented in the mind of a student or physician. The hypothetical patterns of knowledge activation in diagnostic reasoning by a beginning student, an intermediate student and an expert physician will be illustrated on the basis of a case of portal hypertension (Table 6.2). The concept of portal hypertension describes a pathophysiological state and can be considered a typical example of an encapsulating concept: It refers to a limited set of anatomical and causal pathophysiological knowledge including the causes of disease and the resulting clinical consequences. An explanation of portal hypertension is provided in Appendix 6.1, and a schematic representation of this explanation in a canonical model is provided in Appendix 6.2.

The activated knowledge network of a beginning student has been depicted in Figure 6.1. This student activates all kinds of knowledge in studying the case but cannot unequivocally interpret the complaints and physical examination findings of the patient. The knowledge activated does not form one coherent network that points towards one diagnostic hypothesis, but consists of separate pieces of knowledge that are only weakly interrelated. At least four hypotheses are seriously considered by this student. These are constipation, tumor in the abdomen, leukemia, and liver disease. However, none of them can explain all the findings. For this student, the construction of the knowledge network was not only a matter of effortless activation of

Table 6.2. Clinical case of portal hypertension.

A 55-year-old man visiting a family physician complains about piles that are really painful and incidentally bleed. Furthermore, he has a bloated and tense feeling in his stomach which is more bulgy than usual. The last few weeks he had to loosen his belt. He can see his blood vessels around his navel.

The physical examination reveals that the patient has a tensely distended abdomen with an everted umbilicus. Radiating from the umbilicus, tortuous, swollen veins are clearly visible. Percussion of the abdomen reveals that the patient has flank dullness which shifts with changes in position of the patient. A fluid wave is palpable. The spleen is palpable. In rectal examination multiple hemorrhoids are observed, with no further abnormalities.

associated concepts. Active inferencing with the activated biomedical and clinical concepts was also required to construct his representation of the clinical case. The incoherence of this representation reflects the student's lack of knowledge about the disease from which the patient suffers.

Figure 6.2 depicts the activated knowledge network of an intermediate student. This student has understood the relationship between the clinical findings and concludes that the patient has hemorrhoids, caput medusae, an enlarged spleen and ascites. These findings are all strongly associated with the diagnostic hypothesis of liver disease, and, therefore, the student suggests that the patient has a problem with his liver. Subsequently, the student reasons through his knowledge network constructing a pathophysiological explanation that supports his hypothesis. The concept of portal hypertension already plays an important role in the reasoning of this student, but does not yet encapsulate more detailed pathophysiological knowledge: The concept of portal hypertension is not directly linked to signs and symptoms, and is as strongly linked to more detailed pathophysiological concepts as to the concept of liver cirrhosis.

Finally, the knowledge activated by an expert physician has been depicted in Figure 6.3. In studying the case, the expert immediately activates the concepts of hemorrhoids, caput medusae, enlarged spleen, and ascites. Because these concepts are all strongly associated with the concept of portal hypertension, this concept becomes highly activated, and pops up as the first diagnostic hypothesis in the mind of the physician. Subsequently, the possible causes of the problem, as well as one of its serious complications, become highly activated. The expert, therefore, suggests that the patient might have liver cirrhosis (the most common cause of portal hypertension) and need to be examined for esophageal varices, as well as for other signs of liver cirrhosis. Detailed biomedical concepts that are associated with the patient's problem are also activated, but because the relative strength of these associations is lower, their activation level is lower. Thus, the concept of portal hypertension coherently and sufficiently describes the patient's problem, and indeed encapsulates more detailed pathophysiological mechanisms of disease.

Although the networks of knowledge presented do accurately represent which knowledge is activated in case processing by subjects of different levels of expertise, they only roughly represent the time course of the diagnostic process. Whereas students often have to consciously construct their representation of a clinical case needing time and effort to make inferences, experts immediately activate the relevant concepts in diagnosing a routine case. This is assumed to be due to high association strengths between relevant concepts in experts' knowledge networks, and to a more functional differentiation of association strengths between concepts. This structure of expert medical knowledge, thus, explains the automatic pattern recognition character of expert diagnostic reasoning which has been described by many authors (e.g., Norman, et al., 1989; Norman, Tugwell, Feightner & Jacoby, 1985). Furthermore, it explains the finding in Chapter 3, that, if more time is spent on case processing, experts also activate more detailed pathophysiological concepts.

Non-routine diagnosis and explanation by experts

In diagnosing non-routine complex and atypical clinical cases, experts do not automatically form a coherent representation of the patient's problem. Like the beginning student in Figure 6.1, experts may activate several competing diagnostic hypotheses, none of which can adequately explain the signs and symptoms of the patient. Experts then need to elaborate on their knowledge to solve the problem. In other cases, the findings may consistently point towards an encapsulating concept, but not to its most common cause. For example, the concept of portal hypertension may be strongly activated, but the concept of liver cirrhosis may both be activated and inhibited. Because other causes of portal hypertension are not so often seen, and, hence, are not so well tuned, it may cost experts more conscious effort to differentiate between them. In any case, explicit reasoning through the knowledge network takes time. Therefore, non-routine diagnosis is a slower process than routine diagnosis.

In explaining a clinical concept or problem to students and patients, furthermore, experts put time and effort to activate the detailed network of biomedical knowledge that describes the underlying mechanism of disease. Explanation is a constructive process, in which all relevant problem elements become consciously represented in the mind of a physician. Because experts' knowledge networks relating the causes and consequences of disease are so coherently organized, they can easily construct a stable and coherent representation of the concept. This was clearly suggested by the study in Chapter 4 showing that expert physicians have fluent access to a large body of well organized and accurate knowledge.

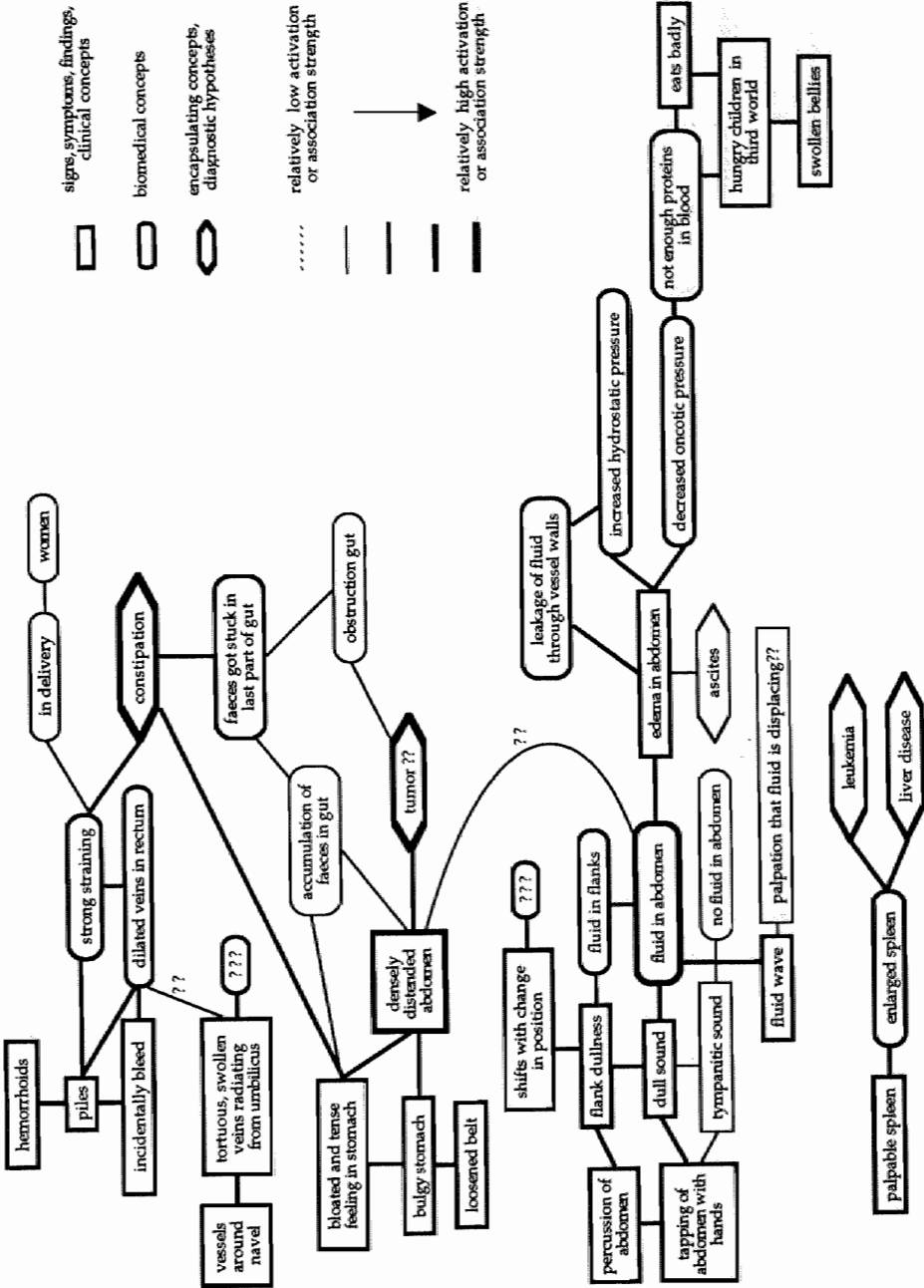


Figure 6.1. Hypothetical activated knowledge network of a beginning student.

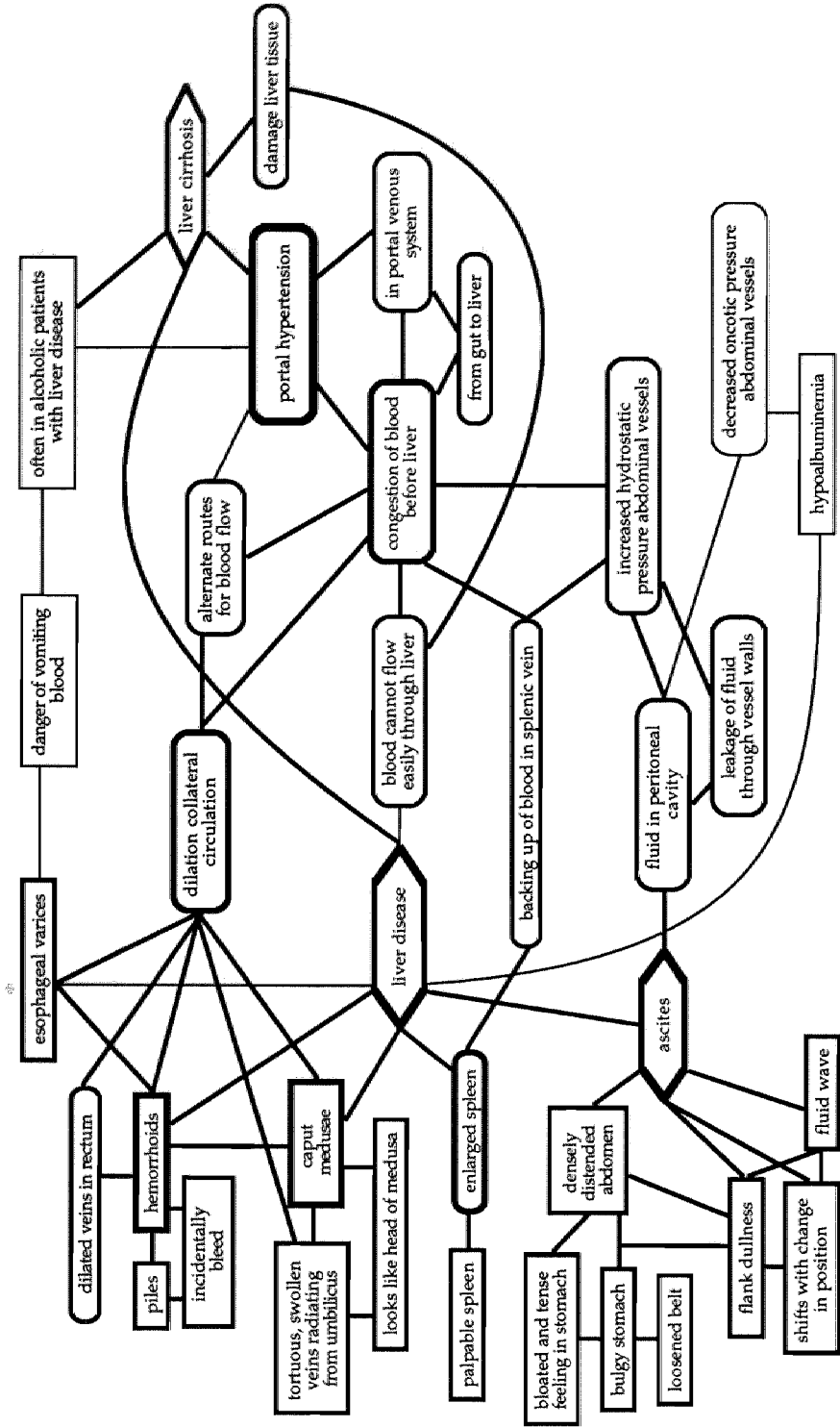


Figure 6.2. Hypothetical activated knowledge network of an intermediate student (see legends of Figure 6.1).

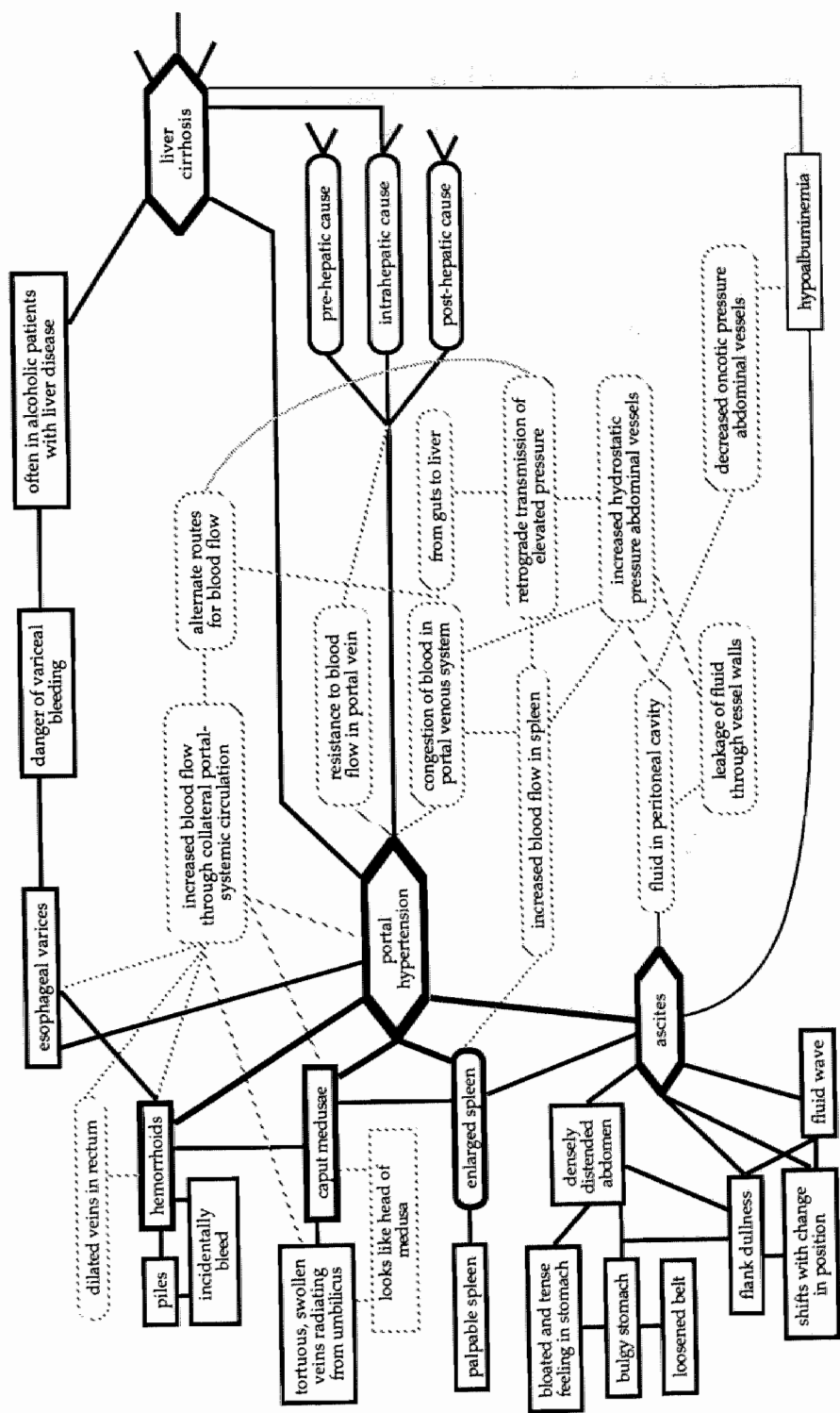


Figure 6.3. Hypothetical activated knowledge network of an expert physician (see legends of Figure 6.1).

Future research

Further research is needed to test the proposed network structure of medical knowledge. A useful method may be the priming paradigm in which the effect of presented information on access of associated information is tested (Anderson, 1983; Meyer & Schvaneveldt, 1971). It has been assumed that the stronger the association between concepts, the more activation spreads to the associated concept, and, therefore, the shorter the response time will be if the concept has to be used in a decision task. If the associations in the knowledge network of students are indeed weaker than those of expert physicians, it may be expected that experts have faster response times than students. Furthermore, by testing many pairs of concepts in a specific knowledge domain, the underlying knowledge network could be represented using techniques such as Pathfinder (Cooke, Durso & Schvaneveldt, 1986; Cooke & Schvaneveldt, 1988; Reitman Olson & Biolsi, 1991) filling up the proximity matrix with response times instead of similarity ratings. Another variant of the priming paradigm that may be used is measuring the response times to a set of biomedical and clinical concepts after subjects diagnosed a clinical case.

Future research should also focus on the learning processes involved in knowledge encapsulation. It has been assumed that repeated use of knowledge in diagnosing and treating patients would induce the encapsulation of elaborate networks of biomedical knowledge into more comprehensive and relevant clinical concepts. This hypothesis could be tested by requesting advanced students to reason through a large set of cases on the same clinical problem while thinking aloud. If knowledge encapsulation is indeed a result of repeated reasoning with knowledge underlying processes of disease, it may be expected that in diagnosing the cases students would gradually use less detailed biomedical knowledge and more clinical encapsulating concepts. Students would recognize the different elements of a patient's clinical problem faster and more easily combine these elements into a diagnostic conclusion. Therefore, the time used for diagnosis can be expected to decrease as a function of the number of clinical cases that are reasoned through. However, if knowledge encapsulation is not so much a gradual process, but rather a matter of insight into the disease processes underlying a set of clinical problems, it can be expected that the shift from biomedical to clinical knowledge use is more abrupt. The time needed for diagnosing a case, then, would show a clear drop after insight has been acquired. What is essential to making a repeated reasoning experiment a good instrument to measure knowledge encapsulation is careful selection of the case material and the form of feedback to students.

Another important point for future research is the generalizability of the process of knowledge encapsulation to other domains than medicine. Knowledge encapsulation, in fact, refers to the elaboration and differentia-

tion of concepts and their interrelations in an associative knowledge net resulting in functional units of knowledge. Since these processes are highly general, it may be assumed that knowledge encapsulation has the potential of being very general. There is indeed evidence from other domains that functional units of knowledge are used in representing and solving problems. For example, in the domain of physics, experts were found to represent problems on the basis of the physics laws underlying the solution of a problem, whereas novices represent these problems on their literal features (Chi, Feltovich & Glaser, 1981; Elio & Scharf, 1990). Through immediate classification of problems by experts, and direct association of category labels to problem-solution knowledge, experts are able to quickly solve a problem. Another example comes from the domain of computer programming, in which experts can represent and think about a program using higher level concepts that refer to the function and programming details of objects, procedures or functions (Anderson, 1990). However, what might make the medical domain essentially different from other domains is the causal nature of the encapsulated knowledge. Knowledge encapsulation as proposed in the medical domain captures not only the categorization of features of disease (e.g., Anderson, 1996; Gluck & Bower, 1988), nor only the detailed pathophysiological knowledge that is summarized, but also the causal relations between clinical features and underlying pathophysiological knowledge. This characteristic of encapsulations makes them particularly strong and coherent units in memory.

Implications for medical education

The process of knowledge encapsulation as outlined in this thesis emphasizes two aspects that are especially important in the learning of effective encapsulations. These aspects are meaningful integration of biomedical and clinical knowledge, and tuning of acquired knowledge to its use in medical practice. An important impetus for both of these processes is the confrontation with real patients in medical training. Students, then, hear patients telling about their problems, see what they look like, and perform physical examinations. They need to interpret these data and formulate differential diagnostic conclusions. To that end, students must relate their experiences to the signs and symptoms described in textbooks and lectures, and may attempt to understand sets of findings in terms of underlying disease processes. Thus, through the interaction with patients, students are stimulated to integrate experiential knowledge to theoretical knowledge leading to more coherent networks of biomedical and clinical knowledge. In addition, when students are confronted with similar clinical problems in different patients, their knowledge becomes tuned in such a way that the most representative

and clinically relevant findings become most strongly associated with higher level clinical concepts or diagnostic labels.

Since medical educators generally assume that a critical mass of knowledge is required before it is useful and ethical to expose patients to students, most medical curricula consist of a preclinical and a clinical period. The preclinical period usually consists of basic sciences, such as anatomy, physiology, molecular biology, immunology and pathophysiology, aiming at a good understanding of the normal and pathological functioning of the human body. The clinical period consists of clerkships at the different clinical units in a hospital directly involving students in the examination and treatment of patients. Many medical schools already provide preclinical programs that aim at the integration of the knowledge from different disciplines by teaching them not only in sequential courses, but also in the context of higher order themes (e.g., the gastro-intestinal tract; senses, neurons and muscles), or types of clinical problems (e.g., abdominal complaints; fatigue and weight loss). Also the training of skills, such as history taking and techniques for physical examination, have already been included in many preclinical programs. However, the transition from student to clerk seems still too abrupt (e.g., Boshuizen, 1996; Boshuizen & Schmidt, 1992; Patel et al., 1989a; Patel & Kaufman, 1995). Even in a problem-based curriculum in which students acquire and apply knowledge to solve and understand written descriptions of clinical problems, they cannot easily apply this knowledge in dealing with real patients in the clerkships. Skills training, furthermore, may focus too much on the learning of the skill itself rather than on the content of the interaction with a patient or on the rationale of the technique. To bridge the gap between theoretical and experiential knowledge, therefore, it seems useful to merge the preclinical and clinical program: Medical students should already be given the opportunity to interact with real patients and their problems before they enter the clinic as clerks, while clerks should be given the opportunity to discuss and reflect on the patient problems they encountered in the clinic in terms of the underlying pathophysiology.

A possible gap-bridging workform in the "pre-clerkship" period is the introduction of real patients in tutorial groups. Clinicians should supervise these small group sessions and guide students in interviewing and examining patients, in interpreting the results, and in formulating a well-founded diagnostic conclusion based on a thorough explanation of the underlying pathophysiological processes of disease. In these sessions, students will not only see, hear, feel and smell the signs and symptoms of disease as they manifest themselves in patients, they will also be stimulated to integrate the different aspects of medical knowledge. As in the small group sessions of problem-based learning, it is important that supervising clinicians display *cognitive congruence* with students in their teaching (Moust, 1993; Schmidt & Moust, 1995). This means that teachers must seriously attempt to explore and understand the problems students have with the subject matter, as well

as to explain these problems in terms that are comprehensible for them. After such a real patient tutorial group session, students might be extra motivated to deepen their knowledge of the clinical problem by studying the literature.

Tutorial group sessions should be continued in the clinical clerkships to discuss students' experiences in the clinic, and to explain the clinical problems they encountered in terms of the underlying pathophysiology. Time for self-study should also be incorporated in the clerkship program, because the integration of theoretical and experiential knowledge will be most effective when problems that are just observed in patients are elaborated upon. Just as in tutorial groups, in one-to-one student-teacher interactions in the clinic too, teachers should pay attention to what students actually know and need to learn. Clinicians seem not always realistic in their estimate of students' knowledge level: A study by Scherpbier (1997) showed that supervising clinicians in the clerkships expected nearly the same knowledge and skills from students that were entering a clinical clerkship as from those that had completed a clinical clerkship.

The patients initially selected for small group teaching should have common or typical problems to improve the tuning of knowledge to its practical use. Setting the stage with this kind of cases provides a framework that enables students to easily fit in later experiences (Neame, 1984; Schmidt et al., 1990). First, many cases seen in later practice will coincide with the presented patients and, therefore, improve students' recognition of a disease by its most representative set of signs and symptoms. Second, subtle variations in the manifestations of disease can be related to the more typical ones leading to refinement of the knowledge base. Third, links with other, but differential diagnostic relevant diseases can be made.

A prerequisite for meaningful integration of knowledge is an active attitude of students aiming at the understanding of the subject matter. To guide their study behavior students need to be aware of what they do know, what they roughly know, and what they do not know. In addition, they must be able to estimate the relevance of the knowledge they lack, as well as the level of detail that is required. Therefore, they need useful feedback. This can be provided by teachers or tutors, other students and formal assessments. Feedback by others is also necessary to correct persistent misconceptions they are not aware of. However, students themselves are the ones that continuously need to monitor the adequacy of their knowledge. The study in Chapter 4 has shown that the explanation of clinical concepts to themselves and others is a valuable tool for self-evaluation. Students immediately experienced the limits of their knowledge when they explained the concepts, and frequently indicated what they did not quite understand and what they needed to study more deeply. Moreover, self-explanation improves learning and understanding (e.g., Chi, Bassok, Lewis, Reimann & Glaser, 1989; Chi, De Leeuw, Chiu & Lavancher, 1994). Explanation is a constructive process in

which the integration of knowledge is stimulated and in which conflicting conceptions must be resolved. As a result, the associations between concepts in the knowledge network become stronger and more accurate.

It may be expected that a combination of practical experience with real patients, tutorial groups and self-study will lead to a better and smoother integration of knowledge, so that more coherent networks of knowledge are associated with relevant clinical concepts. Such an integrated curriculum may also be expected to overcome students' experience that medical training and clinical practice are two worlds apart requiring different, hardly compatible types of knowledge (Patel et al., 1989a; Patel, Evans & Groen, 1989b; Patel & Kaufman, 1995).

References

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1990). *Cognitive psychology and its implications*. (3rd ed.). New York: W. H. Freeman and Company.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51(4), 355-365.
- Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1981). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical & Investigative Medicine*, 5, 49-55.
- Boshuizen, H. P. A. (1989). *De ontwikkeling van medische expertise: een cognitief-psychologische benadering* [The development of medical expertise; a cognitive psychological approach]. Doctoral dissertation. Haarlem: Thesis publishers.
- Boshuizen, H. P. A. (1996). *The shock of practice; effects on clinical reasoning*. Paper presented at the Annual Meeting of the American Educational Research Association. New York, NY, April 8-12.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, 33(1), 1-16.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-477.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education*, 19, 61-67.
- Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12, 538-549.
- Cooke, N. M., & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man-Machine Studies*, 29, 407-427.
- Custers, E. J. F. M. (1995). *The development and function of illness scripts. Studies on the structure of medical diagnostic knowledge*. Doctoral Dissertation. Maastricht: Universitaire Pers Maastricht.

- Custers, E. J. F. M., Regehr, G., & Norman, G. R. (1996). Mental representations of medical diagnostic knowledge: A review. *Academic Medicine*, 71, S55-S61.
- De Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). The Hague: Mouton Publishers.
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, 7(2), 149-158.
- Elio, R., & Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science*, 14, 579-639.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving*. Cambridge, MA: Harvard University Press.
- Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt & M. L. De Volder (Eds.), *Tutorials in problem-based learning. New directions in training for the health professions*, (pp. 128-142). Assen/Maastricht: Van Gorcum.
- Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade*, (pp. 275-319). Reading, MA: Addison-Wesley Publishing Company.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117(3), 227-247.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Hobus, P. P. M. (1994). *Expertise van huisartsen. Praktijkvoering, kennis en diagnostische hypothesevorming* [Family physicians' expertise: Practical experience, knowledge and formation of diagnostic hypotheses]. Doctoral dissertation. Maastricht University, the Netherlands.
- Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses: Expert-novice differences. *Medical Education*, 21, 471-476.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory 'Essays in honor of Gordon Bower'*, (pp. 31-60). San Francisco, New York: W. H. Freeman and Company.
- Marre, J. M. J. (1991). Leren in neurale netwerken. *Informatie*, 33(6), 376-388.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Moust, J. H. C. (1993). *De rol van tutoren in probleemgestuurd onderwijs: Contrasten tussen student- en docent-tutoren* [On the role of tutors in problem-based learning: Contrasting student-guided with staff-guided tutorials]. Doctoral Dissertation. Maastricht: Universitaire Pers Maastricht.
- Neame, R. L. B. (1984). Problem-centered learning in medical education: The role of context in the development of process skills. In H. G. Schmidt & M. L. De Volder (Eds.), *Tutorials in problem-based learning. New directions in training for the health professions*, (pp. 33-47). Assen/Maastricht: Van Gorcum.
- Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: A cross-sectional and longitudinal analysis. *Medical Education*, 15, 315-322.
- Norcini, J., & Dawson-Saunders, B. (1994). Issues in recertification in North America. In D. Newble, B. Jolly, & R. Wakeford (Eds.), *The certification and recertification of doctors: Issues in the assessment of clinical competence*. Cambridge: Cambridge University Press.
- Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1166-1174.
- Norman, G. R., Jacoby, L. L., Feightner, J. W., & Campbell, E. J. M. (1979). Clinical experience and the structure of memory. *Proceedings of the 18th Conference on Research in Medical Education*, 214-219.

- Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E. K. M. (1994). Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine*, 6(2), 114-120.
- Norman, G. R., Tugwell, P., Feightner, L. J., & Jacoby, L. L. (1985). Knowledge and clinical problem-solving. *Medical Education*, 19, 344-356.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989a). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling*, (pp. 53-112). Cambridge, MA: The MIT Press.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989b). Reconciling basic science and clinical reasoning. *Teaching and Learning in Medicine*, 1(3).
- Patel, V. L., Evans, D. A., & Kaufman, D. R. (1990). Reasoning strategies and the use of biomedical knowledge by medical students. *Medical Education*, 24, 129-136.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*, (pp. 93-125). Cambridge: Cambridge University Press.
- Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition*, 18(4), 394-406.
- Patel, V. L., & Kaufman, D. R. (1995). Clinical reasoning and biomedical knowledge: Implications for teaching. In J. Higss & M. Jones (Eds.), *Clinical reasoning in the health professions*, (pp. 117-128). Oxford, UK: Butterworth Heinemann.
- Patel, V. L., & Medley-Mark, V. (1986). Relationship between representation of textual information and underlying problem representation in medicine. *Cognitive research series*. Montreal: McGill University, Centre for Medical Education. (CME86-CS1)
- Pollemans, M. (1994). *Kennistoetsing bij huisartsen* [Testing of knowledge in family physicians]. Doctoral dissertation. Maastricht: Universitaire Pers Maastricht.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Massachusetts: MIT Press.
- Reitman Olson, J., & Biolsi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*, (pp. 240-285). Cambridge: Cambridge University Press.
- Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: Three modes of learning. In R. Klatzky & J. W. Cotton (Eds.), *Semantic factors in cognition*, (pp. 37-60). Hillsdale, NJ: Erlbaum.
- Scherpbier, A. J. J. A. (1997). *Kwaliteiten van vaardigheidsonderwijs onderzocht* [Qualities of skills training investigated]. Doctoral dissertation. Maastricht: Universitaire Pers Maastricht.
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-282). New York, NY: Springer Verlag.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), 205-221.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, 21(3), 338-351.
- Schmidt, H. G., Boshuizen, H. P. A., & Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 139-145.
- Schmidt, H. G., & Moust, J. H. C. (1995). What makes a tutor effective? A structural-equations modeling approach to learning in problem-based curricula. *Academic Medicine*, 70(8), 708-714.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, 65(10).
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text-processing of domain-related information of individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 275-290.

- Van Leeuwen, Y. D., Mol, S. S. L., Pollemans, M. C., Drop, M. J., Grol, R., & van der Vleuten, C. P. M. (1995). Change in knowledge of general practitioners during their professional career. *Family Practice*, 12, 313-317.
- Yekovich, F. R., & Walker, C. H. (1987). The activation and use of scripted knowledge in reading about routine activities. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading*, (pp. 145-176). Hillsdale, NJ: LEA, publishers.

Appendix 6.1

Explanation of portal hypertension

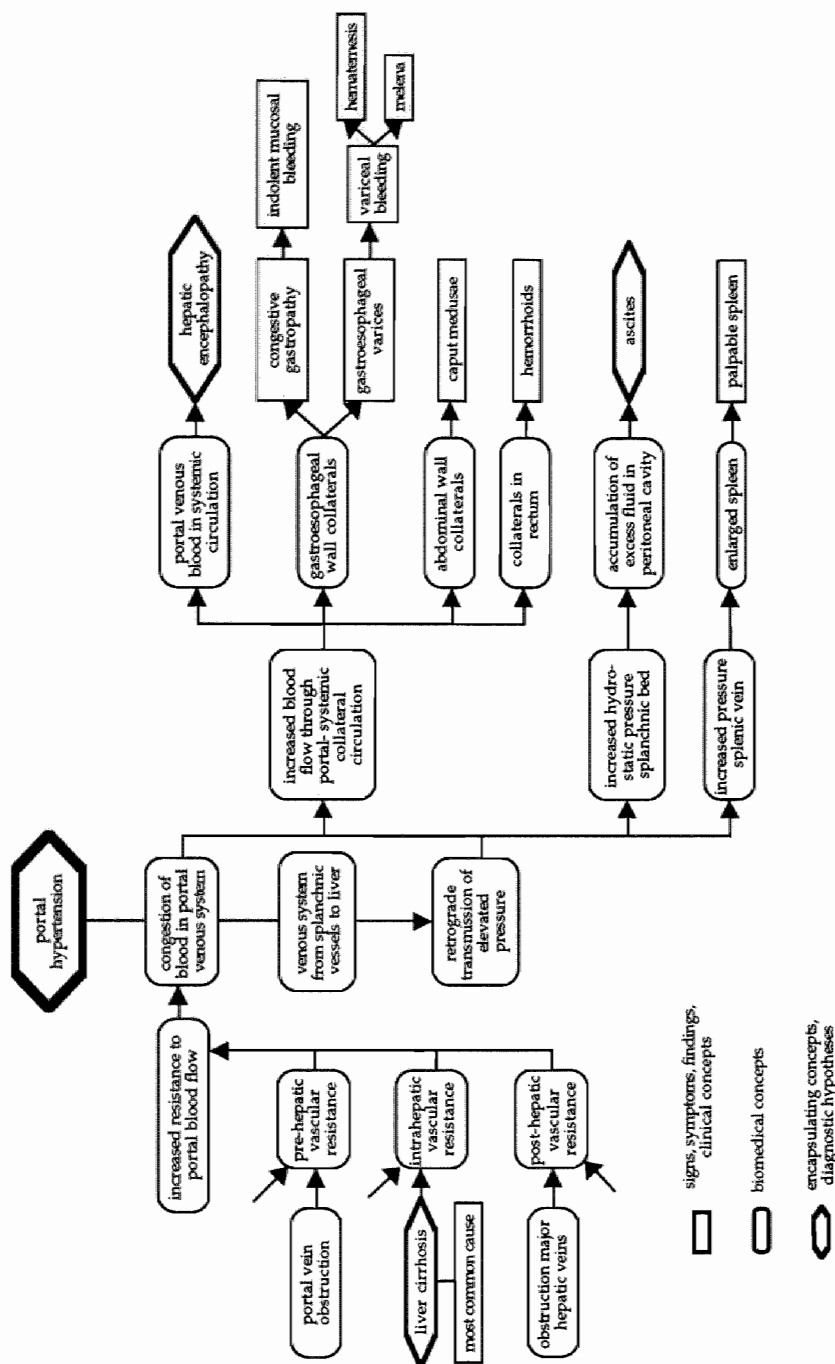
Portal hypertension refers to an elevated blood pressure in the portal venous system, the diseases and pathophysiological mechanisms causing this elevated pressure, and the resulting signs and symptoms to be found in patients. The portal venous system is the venous system from the splanchnic vessels to the liver. It transports absorbed materials from the gut directly to the liver. The materials are processed in the liver and the blood is drained away through the hepatic vein. Any obstruction of blood flow in the portal vein, the liver or the hepatic vein leads to the congestion of blood in the portal venous system.

The obstruction of blood flow in the portal and hepatic veins can for example be caused by a tumor compressing these veins. Most commonly portal hypertension is caused by liver cirrhosis, a disease process in which liver tissue is damaged.

The manifestations of portal hypertension are related to the higher blood pressures developed in the veins connected to the portal venous system. Increased blood pressure in the splenic vein leads to an enlarged spleen which is palpable in physical examination. Increased hydrostatic pressure in the splanchnic capillary bed leads to the leakage of fluid from the capillaries to the peritoneal (abdominal) cavity, and the accumulation of fluid in the peritoneal cavity is called ascites. Furthermore, the portal venous system is connected to the systemic circulation by portal-systemic collaterals, and as a consequence of the high pressure in the portal venous system the blood flow through these collaterals increases. Clinically the collaterals can be detected as gastroesophageal varices which can easily burst and bleed; as dilated veins on the abdomen (caput medusae); and as dilated veins around the rectum (hemorrhoids). The collaterals by-pass the liver so that blood that is not detoxified in the liver enters the systemic circulation and can affect the brain. This may result in a complex neuropsychiatric syndrome called hepatic encephalopathy.

Appendix 6.2

Canonical explanation of portal hypertension



Samenvatting

Kennisencapsulatie

Onderzoek naar de ontwikkeling van medische expertise

Dit proefschrift gaat over de wijze waarop geneeskundestudenten zich na verloop van tijd ontwikkelen tot ervaren artsen. Het uitgangspunt bij het gerapporteerde onderzoek is dat medische expertise het resultaat is van veranderingen in de *aard* en *organisatie* van kennis. Dat wil zeggen dat onder invloed van ervaring kennis zodanig verandert dat experts in staat zijn om snel en doeltreffend een diagnose te stellen bij problemen waarmee zij in de praktijk regelmatig geconfronteerd worden.

In hoofdstuk 1 wordt een onderscheid gemaakt tussen *biomedische* en *klinische* kennis. Biomedische kennis is causaal van karakter en beschrijft het functioneren en disfunctioneren van het menselijk lichaam. Klinische kennis daarentegen is voornamelijk associatief van karakter en omvat kennis over de verschijnselen, classificatie en behandeling van ziekten. Geneeskundestudenten beginnen gewoonlijk met het leren van biomedische kennis voordat zij ingaan op de klinische aspecten van ziekten. In de co-schappen ligt sterk de nadruk op het verwerven van klinische kennis.

Op grond van cognitief psychologisch onderzoek naar de ontwikkeling van medische expertise zijn verschillende standpunten geformuleerd met betrekking tot de rol van biomedische kennis bij het diagnosticeren van klinische problemen. De theorie van *kennisencapsulatie* staat centraal in dit proefschrift. Deze theorie veronderstelt dat gedurende de ontwikkeling van student tot ervaren arts netwerken van biomedische kennis worden samengevat in meer omvattende, klinisch relevante begrippen.

Schmidt en Boshuizen (1992) introduceerden de theorie van kennisencapsulatie om een aantal bevindingen in medisch expertiseonderzoek te verklaren. Hardop-denken studies van diagnostisch redeneren lieten duidelijk zien dat artsen sneller en accurater diagnostische hypothesen formuleren dan studenten. Artsen gebruikten ook veel minder biomedische kennis bij het diagnosticeren van de casus dan studenten. Verder werd herhaalde malen gevonden dat artsen zich minder herinnerden van een bestudeerde casus dan gevorderde studenten, maar meer inferenties maakten op basis van

de casus. Deze bevinding, die bekend is als "the intermediate effect in clinical case recall", duidt erop dat artsen casus op een meer interpreterende en samenvattende manier weergeven dan studenten. Pathofysiologische verklaringen die gegeven werden voor de verschijnselen in een casus lieten tenslotte zien dat de verklaringen van artsen korter en meer samenhangend waren dan die van studenten en veel minder biomedische begrippen bevatten. Schmidt en Boshuizen (1992) veronderstelden dat de snelheid en efficiëntie waarmee ervaren artsen casus representeren en diagnosticeren het resultaat is van directe activatie van *encapsulerende* klinische begrippen. Studenten, daarentegen, zouden alle mogelijke relevante kennis activeren om een medisch probleem op te lossen. Bij kennisencapsulatie zou gedetailleerde biomedische kennis niet verloren gaan. Artsen zouden deze kennis gebruiken in gevallen waarin de klinische kennis niet toereikend is, zoals bij het diagnosticeren van complexe casus en bij uitleg aan studenten.

Vier studies zijn uitgevoerd om de theorie van kennisencapsulatie te toetsen en verder uit te werken. In hoofdstuk 2 wordt een studie beschreven waarin tweede- en vierdejaars geneeskundestudenten en internisten drie casus bestuderen, deze diagnosticeren en achteraf vertellen wat zij zich nog van een casus herinneren. Hierbij werden de instructies van de herinneringstaak en het format van de laboratoriumgegevens gemanipuleerd. Er werd geen verschil aangetoond tussen de proefpersonen die wisten dat ze gevraagd zouden worden de casus te herinneren en degenen die dat niet wisten. Verder maakte het geen verschil of de laboratoriumgegevens in numerieke vorm of in geïnterpreteerde vorm werden gepresenteerd. De resultaten lieten duidelijke expertise-effecten zien. Experts gebruikten minder tijd om de casus te bestuderen en stelden betere diagnoses. Bij de herinneringstaak werd weer een intermediate effect gevonden: internisten noteerden minder elementen uit de casus dan gevorderde studenten, maar maakten meer inferenties die casuselementen samenvatten. Deze resultaten bevestigen de hypothese dat ervaren artsen casus minder uitgebreid, en meer geëncapsuleerd verwerken dan studenten.

In hoofdstuk 3 wordt een studie beschreven waarin de pathofysiologische representatie van casus werd onderzocht bij tweede-, vierde- en zesdejaars geneeskundestudenten en internisten. De proefpersonen kregen vier casus voorgelegd met de vraag de casus te diagnosticeren en een pathofysiologische verklaring te geven voor de symptomen en bevindingen in de casus. Afhankelijk van de experimentele conditie hadden proefpersonen 3 min, 1 min en 15 sec, of 30 sec de tijd om de casus te bestuderen. De verklaringen werden herschreven als semantische netwerken van betekenisvolle concepten. Deze werden vergeleken met modelverklaringen van de casus met betrekking tot zowel de aard van de gebruikte concepten als de relaties tussen de concepten. Zoals verwacht waren de diagnoses en verklaringen van de internisten beter dan die van studenten. De verklaringen van de internisten waren echter minder uitgebreid en minder gedetailleerd. Deze re-

sultaten wijzen erop dat ervaren artsen de pathofysiologische kennis die ten grondslag ligt aan een casus inderdaad op een meer samenvattende manier representeren dan studenten. De kwaliteit van de verklaringen van artsen was bovendien even goed bij alle drie de bestuderingstijden, terwijl de kwaliteit van de verklaringen van studenten verminderde bij verkorting van de bestuderingstijd. In tegenstelling tot wat voorspeld werd op grond van de theorie van kenniscapsulatie (Schmidt en Boshuizen, 1992; 1993b) werd gevonden dat internisten vollediger diagnoses stelden, en langere en meer gedetailleerde verklaringen gaven wanneer zij meer tijd hadden om de casus te bestuderen. Om het effect van bestuderingstijd op de casusrepresentatie van ervaren artsen te verklaren werd verondersteld dat zowel encapsulerende klinische begrippen als gedetailleerde biomedische begrippen deel uitmaken van een associatief kennisnetwerk. De theorie van spreidende activatie kan dan verklaren dat bij grotere tijdsinvestering meer begrippen geactiveerd worden.

De studie in hoofdstuk 4 was niet gericht op het gebruik van encapsulerende begrippen bij het diagnosticeren en representeren van casus, maar op de kennis die wordt samengevat in een encapsulerend begrip. Vierdejaars geneeskundestudenten, co-assistenten en internisten werden gevraagd 20 veel gebruikte klinische begrippen aan de proefleider uit te leggen. Voor elk begrip hadden zij 2 minuten. De uitleg van de proefpersonen werd geanalyseerd op uitgebreidheid, kwaliteit en vloeiendheid. Het bleek dat meer ervaren proefpersonen over het algemeen uitgebreidere, betere en vloeiendere uitleg gaven. Dit wijst erop dat causale pathofysiologische kennis niet verloren gaat bij het opdoen van klinische ervaring, maar juist een samenhangend kennisnetwerk vormt dat gemakkelijk toegankelijk is. Kwalitatieve analyse liet verder zien dat vierdejaars geneeskundestudenten voor sommige begrippen wel een kwalitatief even goede uitleg gaven als internisten, maar duidelijk minder vloeiend en samenhangend. Dit suggereert dat het proces van kenniscapsulatie gradueel verloopt en al optreedt voordat studenten ervaring hebben in de praktijk.

In hoofdstuk 5 wordt verslag gedaan van een replicatieonderzoek van een representatiestudie van Schmidt en Boshuizen (1993b). Tweede-, vierde-, en zesdejaars geneeskundestudenten en internisten kregen vier verschillende casus voorgelegd uit het domein van de interne geneeskunde. De proefpersonen kregen 3 min, 1 min en 15 sec of 30 sec de tijd om een casus te bestuderen. Hun taak was een diagnose te stellen en op te schrijven wat ze zich nog van de casus konden herinneren. Zoals verwacht had expertise een positief effect op diagnostische accuratesse. In tegenstelling tot de bevindingen van Schmidt en Boshuizen werd geen intermediate effect gevonden bij de herinneringstaak, maar een lineair expertise-effect: meer ervaren proefpersonen herinnerden zich meer elementen van de casus. Bovendien werd niet alleen bij studenten maar ook bij internisten een effect van bestuderingstijd gevonden. Deze resultaten wijzen erop dat de in-

ternisten in deze studie de casus op een uitgebreidere en minder geëncapsuleerde manier representeerden dan de internisten in de studie van Schmidt en Boshuizen.

In hoofdstuk 6 volgt na samenvatting van de vier studies een herformulering van de theorie van kennisencapsulatie van Schmidt en Boshuizen (1993a). In hun stadiamodel van expertiseontwikkeling veronderstelden Schmidt en Boshuizen dat kennisencapsulatie een belangrijke rol speelt bij de verandering van de kennisstructuur van causaal netwerk tot zogenaamde ziektescripts. Zij geven echter niet duidelijk aan hoe kennisencapsulatie gerealiseerd wordt en hoe de kennisstructuur van ziektescripts eruit ziet. Het huidige voorstel gaat er vanuit dat expertiseontwikkeling meer een continu proces is, en dat zowel de kennis van studenten als experts georganiseerd is in een netwerk. Als gevolg van onderwijs en ervaring wordt het kennisnetwerk uitgebreid, beter, en meer samenhangend en doelmatig georganiseerd. Kennisencapsulatie verwijst dan naar die veranderingen in het kennisnetwerk waarbij verbindingen tussen encapsulerende begrippen, ziekteverschijnselen en mogelijke oorzaken het sterkst worden. De gedetailleerde netwerken van pathofysiologische kennis worden bij het oplossen van routineproblemen dan als het ware omzeild, terwijl ze bij moeilijkere problemen eenvoudig geactiveerd kunnen worden. Vervolgens komt aan bod hoe de theorie van kennisencapsulatie verder getoetst zou kunnen worden. Tenslotte wordt geconcludeerd dat de integratie van biomedische en klinische kennis meer gestimuleerd zou moeten worden in het medisch onderwijs, en wordt hiervoor een aantal mogelijkheden geopperd.

Curriculum Vitae

Margje van de Wiel was born in 's Hertogenbosch on November 17, 1964. After completing gymnasium B at the Dr. Moller College in Waalwijk in 1983, she studied Biology for two years at the Catholic University of Nijmegen. From 1985 to 1990 she studied Psychology in Nijmegen graduating in Experimental Psychology. Her Master's thesis investigated the role of color in information transfer. After her graduation she developed, as a participant in a postgraduate course in knowledge technology, a computer-aided-instruction program at the National Institute for Educational Measurement (Cito) in Arnhem. In 1991 she started her Ph.D. research on the development of medical expertise at the Department of Educational Development and Research, Maastricht University. In the spring of 1995 she continued her studies at the Centre for Medical Education, McGill University, Montreal, Canada. At Maastricht University she worked as tutor and skills trainer in the problem-based learning curricula of Health Sciences and Psychology. Since 1996 she has been working as a researcher at the Skillslab of the Faculty of Medicine in Maastricht, where she investigates the effects of real patients tutorial groups on the knowledge of students.